INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



LA TEORIA DEL ARCO IRIS

Junio 1977

140 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 9
Junio 1977

INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN

SCIENTIFIC AMERICAN

COMITE DE REDACCION: Gerard Piel (Presidente); Dennis Flanagan, Francis Bello, Philip Morrison; Trudy E. Bell; Brian P. Hayes; Jonathan B. Piel; John Purcell; James T. Rogers; Armand Schwab, Jr.; Jonathan B. Tucker; Joseph Wisnovsky
DIRECCION EDITORIAL: Dennis Flanagan
DIRECCION ARTISTICA: Samuel L. Howard
PRODUCCION: Richard Sasso
DIRECTOR GENERAL: Donald H. Miller, Jr.

INVESTIGACION Y CIENCIA

DIRECTOR: Francisco Gracia Guillén
REDACTOR JEFE: José María Valderas Gallardo
PRODUCCION: Manuel Estrada Herrero
PROMOCION: Pedro Clotas Cierco
EDITA: Prensa Científica, S.A.
Calabria, 235-239
Barcelona-15
ESPAÑA

Colaboradores de este número:

Asesoramiento y traducción:

Gerardo Espinosa: *La importación de gas natural licuado;* Monique Robert: *El interferón;* Manuel Puigcerver: *Teoría del arco iris;* José Luis Amorós: *Biocristales;* Luis Bou García: *Algoritmos;* Miquel Gich: *La colisión entre la India y Eurasia;* Ramón Canal: *Los acompañantes de las estrellas de tipo solar;* Gerardo Espinosa: *Taller y laboratorio;* Luis Bou García: *Juegos matemáticos.*

Libros:

Juan José Acero



LA PORTADA

La fotografía de la portada muestra un arco iris formado en el rocío de una cascada en Bridal Veil Creek, en el Yosemite National Park de California. El brillo y la pureza de sus colores son poco comunes. También se distingue por su escala aparente: se diría que el arco se halla a poca distancia. En realidad, un arco iris no puede tener un emplazamiento definido. Se trata de un cono de rayos de luz cuyo vértice está en el observador (véase "Teoría del arco iris", por H. Moysés Nussenzveig).

Suscripciones:

Prensa Científica, S.A.
Departamento de suscripciones
Calabria, 235-239
Barcelona-15 (España)

Distribución para España:

Distribuciones de Enlace, S.A.
Ausias March, 49
Barcelona-10 (España)

Distribución para los restantes países:

Editorial Labor, S.A.
Calabria, 235-239
Barcelona-15 (España)

Condiciones de suscripción:

España:
Un año (12 números): 1.540 pesetas

Extranjero:
Un año (12 números); 30 U.S. $

Ejemplar atrasado ordinario:
160 pesetas

Ejemplar atrasado extraordinario:
225 pesetas.

PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la cubierta de Richard Rowan, Photo Researchers, Inc.

Página	Fuente
7	General Dynamics Corporation
8-13	Allen Beechel
17	EROS Data Center
18-19	Andrew Tomko
20-21	Alan D. Iselin
22-23	Andrew Tomko
24	Alan D. Iselin
25	EROS Data Center
26	V. P. Solonenko
27	EROS Data Center
31	Ben Rose
32-38	Jerome Kuhl
43-53	Lorelle M. Raboni
54	Douglas Faulkner
55	Shinya Inoué
56	Patricia J. Wynne
57	Shinya Inoué
58	Patricia J. Wynne
59-60	Kayo Okazaki y Kent McDonald
63	Kit Peak National Observatory
64-65	George V. Kelvin
66-67	George V. Kelvin *(dibujos);* Helmut A. Abt, Kitt Peak National Observatory *(espectros)*
68	George V. Kelvin
73	José Egozcue
75	Joel Ito, *Primate News (arriba)* José Egozcue *(abajo)*
76	*Primate News (arriba)* José Egozcue *(abajo)*
77	García y col., *Folia primat.* 25:313 (1976)
78	*Primate News*
79	Joaquín Egozcue *(arriba)* Rubio y col. *(abajo) Folia primat.* 26:306 (1976)
80	José Egozcue y col. *(arriba) Humangenetik* 18:77 (1973) Rubio y col., *Boll. Zool.* 40:399 (1973), *(abajo)*
83	Sharon Beals, Black Star
84-92	Gabor Kiss
96-100	Ilil Arbel

Dep. legal: B. 38.999-76
Fotocomposición: Fort, S.A.
Rosellón, 33 - Barcelona-15
Cayfosa. Santa Perpetua de Moguda. Barcelona.
Printed in Spain - Impreso en España

Cartas

Señor Director:

El artículo de Kosta Tsipis sobre los misiles crucero (*Investigación y Ciencia,* abril, 1977) es revelador e interesante. Y lo es, entre otras cosas, porque indica una política a seguir muy válida en punto a las negociaciones sobre el desarme y el control de armamentos.

La justeza al afirmar que ha llegado el momento de ejercer una "mutua restricción" en lo concerniente al desarrollo y despliegue de misiles crucero de largo alcance es un principio que exige su inmediata incorporación en la prosecución de las conversaciones bilaterales SALT. Me sumo a la opinión defendida por el autor de que conviene una fase inicial de restricción unilateral, restricción que, por razones obvias, debería corresponder primero a Estados Unidos, por hallarse mucho más adelantados en este aspecto que cualquier otro competidor.

Sin embargo, lo que me mueve concretamente a escribir estas líneas es la importancia de recordar a todos los lectores el interés de los demás países, aparte de las dos superpotencias, en el desarrollo de los misiles crucero, habida cuenta de sus posibilidades de revolucionar el sistema de transporte de bombas. Si se construyera un misil crucero para lanzamiento desde tierra o desde buques, con misiones y alcances tácticos, de acuerdo con lo descrito por el doctor Tsipis (es decir, un misil de tamaño pequeño, alta precisión y diseñado para transporte de armas convencionales), no debería ni podría seguir siendo un monopolio de las superpotencias. Su notable valor exigiría que se pusiera a disposición de los demás países; lo cual debiera ser en lo referente tanto al mercado del misil cuanto al de su diseño.

Pero apremia más, al objeto de promover la equidad internacional y la limitación de armamentos, que se internacionalice todo avance en el terreno del reconocimiento y que se logre una mesa de negociaciones multilateral, no exclusivamente bilateral como hasta el presente, según dejé bien sentado en mi libro *The Game of Disarmament.*

ALVA MYRDAL

Ministerio de Asuntos Exteriores
Estocolmo

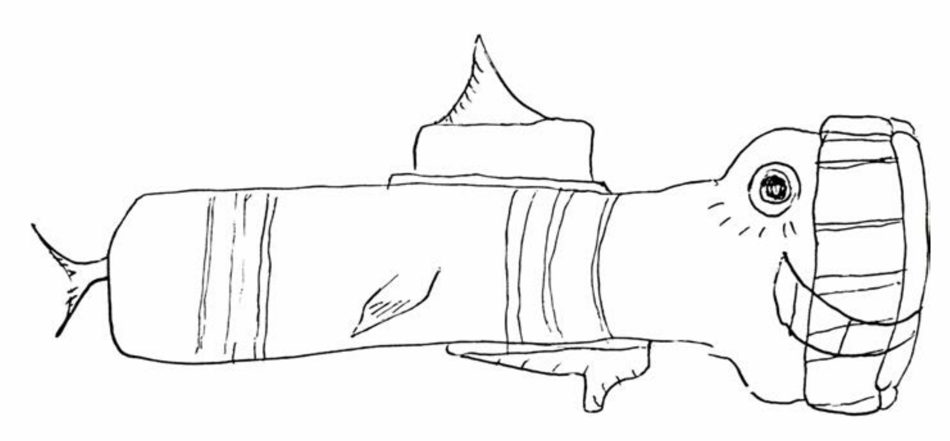
Señor Director:

Como posdata a mi artículo sobre los peces luminosos destellantes (*Investigación y Ciencia,* mayo, 1977) me gustaría relatarle un descubrimiento extraordinario que parece corroborar la teoría de la evolución convergente. En el mes de marzo de 1975 volvía de las Islas Comore con las primeras muestras in vivo del pez luminoso destellante *Photoblepharon* para experimentar con ellas en un acuario. Tras pasar treinta horas sin dormir en aviones y aeropuertos de tres continentes, desembarqué en San Francisco con los peces cuyos ojos (y órganos luminosos) aparecían tan legañosos como lo estaban los míos. Al día siguiente, cuando me dispuse a ordenar los papeles de mi mesa de trabajo, que se habían ido acumulando durante seis semanas de ausencia, hallé, entre la correspondencia, el dibujo que le adjunto de un pez luminoso fantásticamente imaginado por personal colaborador de mi departamento *(véase la superior de las dos ilustraciones inferiores).* Ingenioso, pensé.

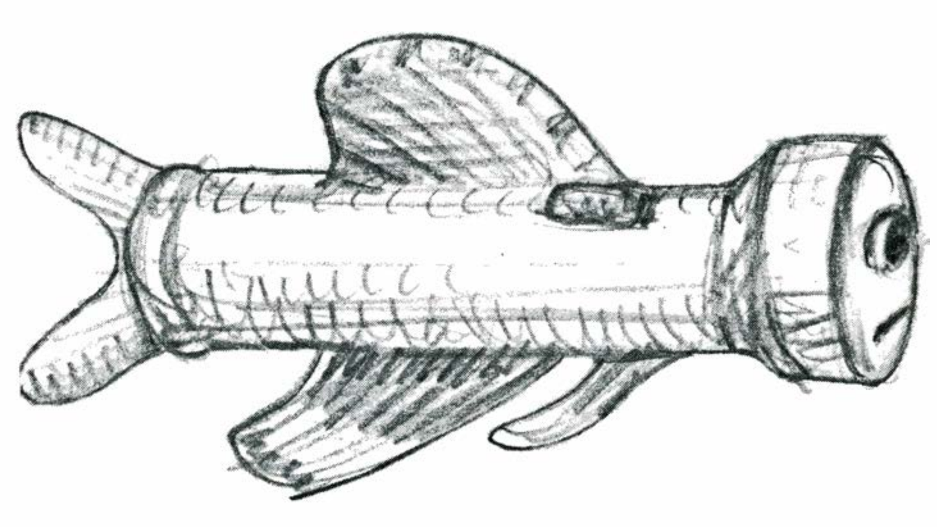
Ahora, a los dos años de aquel esbozo, recibo otro dibujo de un organismo notablemente similar imaginado por Tom Prentiss para ilustrar el artículo de referencia *(véase la ilustración inferior).* Este ejemplo de especiación alimentada por pilas, evolucionada en alopatría en extremos opuestos del Nuevo Mundo (San Francisco y Nueva York), hace que le den a uno ganas de pensarse dos veces los fundamentos marinos (¿pisciformes?) de nuestros principios darwinianos.

JOHN E. MCCOSKER

Director
Steinhart Aquarium
California Academy of Sciences
San Francisco



Pez linterna alimentado por pilas del Steinhart Aquarium, San Francisco



"Photoengendron", género de origen independiente que hace verosímil la teoría de la evolución

Los autores

ELISABETH DRAKE y ROBERT C. REID ("La importación de gas natural licuado") son gerente de la Hazard Assessment and Control Unit de Arthur D. Little, Inc., y director general adjunto del Liquid Natural Research Center, en el Instituto de Tecnología de Massachusetts, respectivamente. Drake se licenció y se doctoró en ingeniería química por el MIT, y trabaja en Arthur D. Little desde 1958. Su larga experiencia en el terreno de la tecnología criogénica, seguridad y materiales peligrosos y metodología de los riesgos le capacitan para el análisis de la seguridad y viabilidad de las industrias dedicadas a las substancias peligrosas (el gas natural entre ellas). Reid se doctoró también en ingeniería química por el MIT, en donde llegó a engrosar su claustro profesoral con el tiempo.

PETER MOLNAR y PAUL TAPPONNIER ("La colisión entre la India y Eurasia") son profesor adjunto del departamento de ciencias planetarias y de la tierra, del Instituto de Tecnología de Massachusetts, y profesor de investigación del Centre National de la Recherche Scientifique (C.N.R.S.) (Francia), en la Universidad de Montpellier, respectivamente. Molnar se licenció en físicas en el Oberlin College, en 1965. Su pasión por el aire libre le indujo, en su etapa de postgraduado, a especializarse en geofísica, en el Lamont-Doherty Geological Observatory, de la Universidad de Columbia. Tras doctorarse por la Universidad de Columbia, en 1970, permaneció en ella otro año y, luego, otros dos más en la Scripps Institution of Oceanography. Desde 1974 ha estado trabajando en el MIT. Tapponnier nació y se educó en la ciudad de Annecy (Alpes franceses) especializándose en física y matemáticas (Lyon). En 1967 ingresó en la École Nationale Supérieure de Mines de París, donde estudió geología y geofísica. En 1971 empezó sus investigaciones tectónicas en Montpellier.

DEREK C. BURKE ("El interferón") es profesor de ciencias biológicas en la Universidad de Warwick. Tras obtener su B. Sc. y doctorarse en química por la Universidad de Birmingham, pasó a la Universidad de Yale en 1953, donde se dedicó al estudio de los nucleósidos de las esponjas del Caribe. Volvió a Inglaterra en 1955 para trabajar con Alick Isaacs, ya fallecido, en el National Institute for Medical Research, sobre el ácido nucleico de los virus de la gripe. Burke empezó a investigar en el campo del interferón escasos meses después del descubrimiento de esta sustancia por Isaacs y Jean Lindenmann en 1957. Desde entonces, el interferón no ha dejado de ser su objeto de estudio, constituyendo algunas veces el único centro de interés en su labor investigadora y hallándose otras veces combinado con investigaciones sobre la bioquímica de los virus. Permaneció nueve años en la Universidad de Aberdeen. Desde 1969 se encuentra en Warwick.

DONALD E. KNUTH ("Algoritmos") es profesor de ciencias de la computación en la Universidad de Stanford. Se graduó en matemáticas en el Case Institute of Technology, doctorándose en 1963 por el Instituto de Tecnología de California. Permaneció en el Cal Tech durante otros cinco años, donde llegó a ser catedrático de matemáticas. En 1968 ingresó en el departamento de ciencias de la computación de la Universidad de Stanford. Knuth ha recibido numerosas distinciones por sus trabajos de investigación, entre los que se cuenta el A.M. Turing Award of the Association for Computing Machinery, en 1974. El citado galardón menciona la serie de siete volúmenes, titulada *The Art of Computer Programming* que Knuth está escribiendo, y de la que espera tener terminados los cinco primeros en 1980.

SHINYA INOUÉ y KAYO OKAZAKI ("Biocristales") son profesor de biología y director del programa de citología biofísica en la Universidad de Pennsylvania, y profesor adjunto de biología de la Universidad Metropolitana de Tokio, respectivamente. Ambos recibieron su preparación en investigación en la Estacion Biológica Marina Misaki, en el Japón. Inoué se graduó en 1948 en Princeton, residiendo en adelante en los Estados Unidos. Antes de asociarse a la facultad de Pennsylvania, enseñó en la Universidad de Washington, en la Universidad Metropolitana de Tokio y en la Universidad de Rochester, y fue director del departamento de citología y anatomía de la facultad de Medicina de Dartmouth. Hoy en día es fideicomisario del laboratorio de biología marina de Woods Hole, Mass. Okazaki es una de las pocas mujeres que tienen el rango de profesor adjunto en el Japón. Su investigación más importante se ha centrado en la embriología del erizo de mar.

HELMUT A. ABT ("Los acompañantes de las estrellas de tipo solar") trabaja como astrónomo en el Observatorio Nacional de Kitt Peak, en Arizona. Tras licenciarse en la Northwestern University obtuvo su doctorado en 1952 en el Instituto de Tecnología de California. Pasó un año investigando en el Observatorio de Lick, de la Universidad de California, y estuvo desde 1953 hasta 1963 en el Observatorio Yerkes, de la Universidad de Chicago. En 1955 y 1956, dedicó gran parte de su trabajo de campo a la búsqueda de un emplazamiento para el actual observatorio de Kitt Peak; reside en Kitt Peak desde 1963.

JOSE EGOZCUE ("Evolución cromosómica de los primates") es profesor agregado de biología celular de la Universidad Autónoma de Barcelona. Cursó la carrera de medicina en esa misma ciudad, obteniendo el grado de doctor por la Universidad de Barcelona en 1969. En 1965 fue nombrado director del departamento de genética del Oregon Regional Primate Research Center, donde permaneció hasta 1969 dedicado al estudio de los cromosomas de los primates. En 1969 fue nombrado director del departamento de citogenética del Instituto de Investigaciones Citológicas de Valencia. En 1970 se incorporó a la recién creada Universidad Autónoma de Barcelona como profesor agregado, formando también parte del Instituto de Biología Fundamental de esta misma Universidad, donde dirige el laboratorio de citogenética.

H. MOYSÉS NUSSENZVEIG ("Teoría del arco iris") es profesor de física en la Universidad de São Paulo, Brasil. Nacido en São Paulo en 1933, se graduó en física en aquella universidad en 1954, y se doctoró por la misma institución en 1957. Después, se trasladó al Centro Brasileño de Investigación Física en Río de Janeiro, donde fue nombrado profesor titular en 1962. El año siguiente marchó a los Estados Unidos. Tras un curso como miembro visitante del Courant Institute of Mathematical Sciences en la Universidad de Nueva York y otro año en el Institute for Advanced Study en Princeton, entró a formar parte del claustro de la Universidad de Rochester. Allí permaneció hasta 1975, año en que regresó a Brasil, a su São Paulo natal.

La importación de gas natural licuado

En unos países escasea el gas natural y sobra en otros. Una solución para lograr que este importante combustible llegue a los centros de consumo sería transportarlo por vía marítima, pero ¿qué riesgos conlleva?

Elisabeth Drake y Robert C. Reid

El consumo de energía en los Estados Unidos, que llegó a una cota máxima de 18.900 billones de kilocalorías en 1973 y luego disminuyó ligeramente, está aumentando de nuevo; la Federal Energy Administration prevé que alcance los 21.420 billones en 1980 y los 23.940 billones de kilocalorías en 1985. Alrededor del 30 por ciento de la energía proviene del gas; los incentivos para mantener esa proporción conforme crece la demanda de energía son poderosos. No obstante, la producción nacional de gas natural ha decaído notoriamente desde que alcanzó su punto máximo en 1973; parece probable que disminuya aún más rápidamente en el futuro, a menos que se encuentren y desarrollen nuevas fuentes. Una manera de suplementar los menguantes abastecimientos de gas natural es importar en gran escala gas natural licuado desde zonas de ultramar que lo posean en abundancia y cuyo mercado interno sea pequeño o inexistente. Cumple preguntarse hasta qué punto es válida esta opción, es decir, hasta qué punto los argumentos en pro de la importación de gas licuado pesan más que los argumentos en contra.

El problema inmediato que se presenta al transportar grandes cantidades de gas natural por cualquier medio que no sea una tubería consiste en que, a las temperaturas corrientes, un gas ocupa un gran volumen. En cambio, si el gas natural se enfría hasta −162 grados Celsius, se convierte en un líquido cuyo volumen es 600 veces menor que el del gas. De ese modo, se pueden resolver los problemas relacionados con el transporte del combustible a lugares a los que no puede llegar una tubería y con su almacenamiento en grandes cantidades. En el caso particular del gas natural licuado los problemas principales son: hallar materiales que contengan sin riesgo una sustancia tan fría y determinar qué medidas de seguridad deben adoptarse contra un posible escape del líquido.

Los argumentos en favor de la importación en gran escala de gas natural licuado pueden resumirse rápidamente, pues no son complicados. Los Estados Unidos necesitan más fuentes de energía. El gas resulta muy adecuado como combustible porque se quema limpiamente y se distribuye con facilidad a los consumidores. Muchos hogares, fábricas y edificios públicos tienen hoy instalada la calefacción a gas. El gas natural, gran parte del cual aparece como subproducto del petróleo con el que viene relacionado, se sigue derrochando en muchas zonas: se vierte a la atmósfera o se quema al no disponerse de un mercado local. Se dispone de una tecnología para licuar el gas natural, para transportarlo a través de los océanos y para almacenarlo en tierra. Además, esa tecnología es menos compleja y menos costosa que la tecnología para extraer gas sintético del petróleo o del carbón. Ya existe un importante comercio internacional de gas natural licuado: Japón importa alrededor del 80 por ciento de su gas en forma líquida, Europa occidental un 5 por ciento y los Estados Unidos algo menos del 1 por mil. En realidad, los Estados Unidos actúan hoy como exportadores netos de gas natural licuado: las exportaciones desde Alaska hasta el Japón son alrededor del triple de las importaciones que, procedentes de Argelia, llegan a un terminal en Massachusetts. (El pasado invierno, a causa de la crisis en el abastecimiento de gas, se autorizó por primera vez el embarque de gas natural licuado desde Alaska a Massachusetts, pasando por el Canal de Panamá.) Las cifras correspondientes al año 1976 fueron: 8 billones de kilocalorías exportadas a Japón y 2,7 billones de kilocalorías importadas a Massachusetts. Para exponerlo en unidades más familiares, las importaciones anuales fueron de 450.000 metros cúbicos; expresado por la capacidad de los buques que actualmente transportan gas natural licuado, las importaciones representaron unas doce cargas completas.

Los argumentos en contra de la importación en gran escala de gas natural licuado se centran en las medidas y condiciones de seguridad de semejante empresa. El gas natural licuado puede ser peligroso si se manipula con descuido o si se liberan grandes cantidades del mismo en un accidente contra el cual se haya dispuesto una protección insuficiente, como ocurre con la gasolina y el gas natural. Debe tenerse en cuenta, asimismo, que el gas licuado tiene una temperatura bajísima.

En 1941 se construyó en Cleveland una instalación para licuar gas natural y almacenar el líquido. En 1944 falló uno de los tanques de almacenamiento y se produjo una catástrofe. Alrededor del tanque no se había construido ningún dique que contuviera la totalidad del líquido en el caso de que ocurriera una fuga. El gas licuado se derramó sin obstáculos hacia la zona circundante. El helado líquido hirvió rápidamente, y los vapores pronto llegaron hasta una fuente de ignición, que desencadenó un inmenso incendio en el que murieron 128 personas.

Este accidente hizo retroceder en muchos años a la incipiente industria del gas natural licuado. Pasaron casi dos décadas antes de que se reexaminaran las ventajas de esta forma del combustible. La posibilidad de hacerlo se suscitó en parte a raíz de los trabajos de la NASA y



BUQUE-CISTERNA PARA GAS NATURAL LICUADO en construcción, por la Quincy Shipbuilding Division, de la General Dynamics Corporation. Una grúa baja una de las cinco esferas metálicas, que contendrán el gas natural licuado, hasta una concavidad preparada al efecto en el casco del barco. Cada esfera tiene capacidad para 25.000 metros cúbicos de gas natural licuado y en el buque cabrán cinco, protegida cada una por una estructura hemisférica blanca. Las esferas se construyen por separado y se transportan hasta el astillero en gabarra.

del Department of Defense sobre los combustibles para cohetes; de esos trabajos salieron materiales perfeccionados para el almacenamiento de líquidos criogénicos y nuevas precauciones de seguridad para su manipulación.

En la década de los años 60, varias compañías municipales proyectaron y construyeron instalaciones que utilizaban gas natural licuado en "horas punta", es decir, que suministraban combustible en momentos de demanda desacostumbradamente elevada. Esas compañías compraban, licuaban y almacenaban gas de gaseoducto en períodos de baja demanda y en los que su precio era inferior al normal. Más tarde, durante las rachas de frío, en las que la capacidad del gaseoducto llegaba a su límite y el precio del gas se elevaba, el gas natural licuado podía revaporizarse y distribuirse por las tuberías matrices que servían a los clientes de la compañía. Actualmente hay más de 60 instalaciones de éstas, en servicio, en Estados Unidos y Canadá.

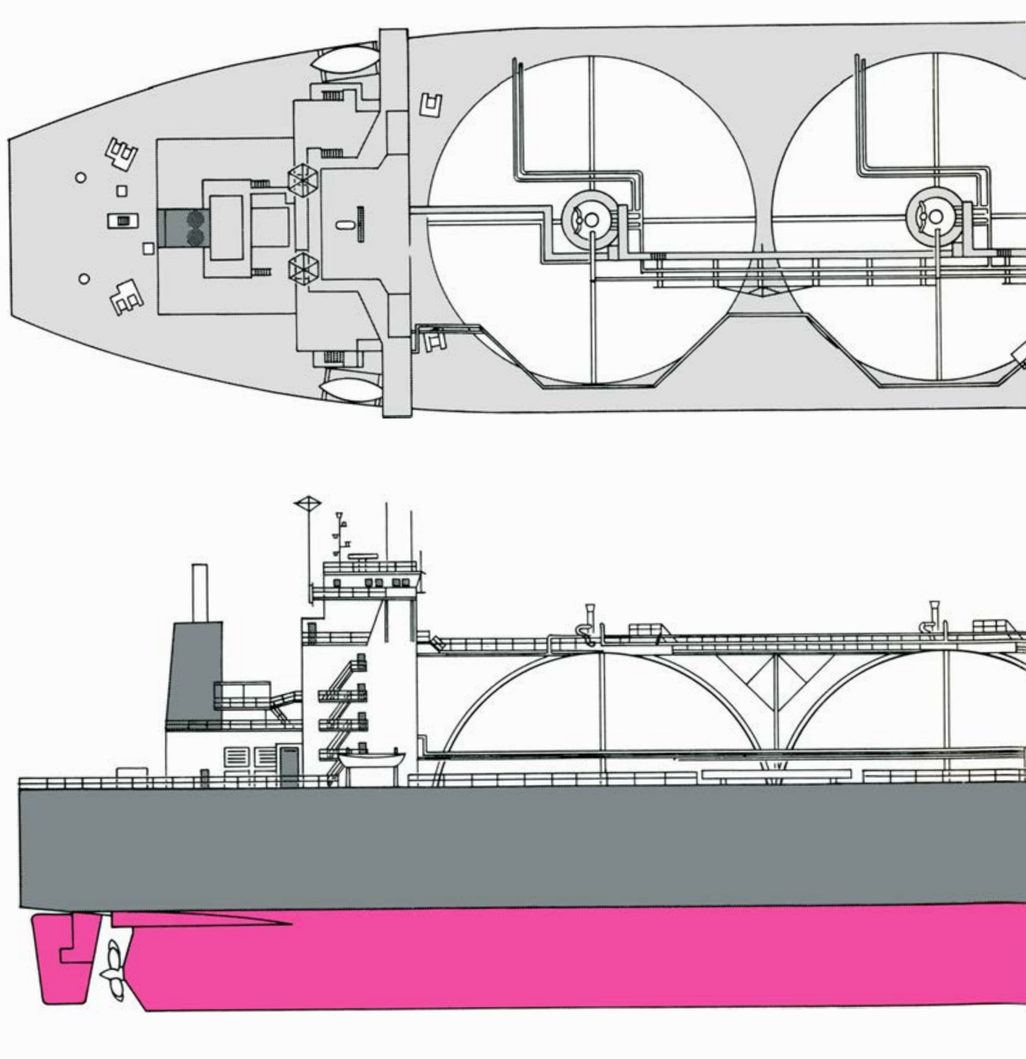
Las instalaciones para "horas punta" han demostrado ser económicamente beneficiosas y muestran un excelente historial de seguridad. Desempeñaron una importante función al mantener el servicio de gas a los hogares durante los períodos insólitamente fríos del pasado invierno. Así pues, a esta escala, el uso de gas natural licuado como combustible parece ser una medida oportuna. Ahora bien, a causa del agotamiento de las reservas de gas en la extensa zona de los 48 estados contiguos de Estados Unidos, está aumentando el incentivo para la importación de gas natural licuado para operaciones de carga fundamental, es decir, de demanda continua de energía. Un terminal destinado a este uso recibiría regularmente gas natural licuado traído por una flota de buques-cisterna, lo almacenaría en grandes tanques y lo revaporizaría continuamente para enviarlo como combustible a través de la red de tuberías.

Los buques-cisternas proyectados para operaciones a esta escala tendrían, cada uno, una capacidad máxima de 165.000 metros cúbicos, lo cual es suficiente para cubrir un campo de fútbol con una capa de líquido de unos 17 metros de profundidad. La capacidad de los tanques del terminal receptor sería de dos a cuatro veces el de la estructura de 50.000 metros cúbicos, típica de las actuales instalaciones usadas para suplementar la demanda en horas punta. Al contemplar estas perspectivas, muchas personas han expresado su preocupación sobre lo que podría ocurrir si un buque-cisterna lleno de gas licuado chocase, si un avión cayese sobre un tanque de almacenamiento, si en una zona en que hubiera instalado un tanque ocurriese un terremoto o si se sabotease una planta o ramal. La cuestión que debe afrontarse es si las disposiciones de seguridad que pudieran adoptarse bastarían para permitir que la importación de gas natural licuado siguiese adelante con unos riesgos soportables para el público.

Un motivo que justifica la preocupación por el gas natural licuado es el accidente que ocurrió en 1967 en una instalación de almacenamiento situada en Staten Island, que es parte de la ciudad de Nueva York. El accidente afectó a un tanque que había servido de almacén de gas natural licuado. De fuera adentro, el tanque estaba hecho de una pared de hormigón, una capa aislante de espuma de poliuretano y de un recipiente para el líquido. El recipiente era una membrana hecha de película aluminizada del plástico Mylar.

A los dos años de haber entrado en servicio el tanque, se produjo una fuga en la película de Mylar. Se abandonó dicho depósito, y el gas natural licuado que contenía se bombeó y se vaporizó para consumo local. Se insufló luego nitrógeno gaseoso en el tanque para caldearlo y para eliminar cualquier residuo de vapor inflamable. Al cabo de un tiempo se dejó entrar aire en el tanque y penetraron obreros para repararlo.

Mientras estaban los obreros en su interior, se produjo un incendio en la capa aislante. Los gases calientes de la combustión elevaron la presión interna. El techo se elevó, tal como estaba proyectado para hacerlo en caso de crearse una presión excesiva dentro del tanque; al asentarse, se derrumbó hacia el interior, matando a 40 obreros. Aunque el accidente nada tuvo que ver con el gas natural licuado ni puso en peligro la seguridad pública, suscitó la oposición local a un terminal que se estaba construyendo en las cercanías para la impor-



BUQUE ESPECIALIZADO del tipo que está construyendo la General Dynamics Corporation. Está proyectado para transportar 125.000 metros cúbicos de gas natural licuado en cinco tanques

tación de gas licuado. Otro motivo de preocupación procedía de que el terminal implicaba la llegada de buques con carga de gas natural licuado en aguas en las que se habían producido accidentes. Permanece sin resolver la cuestión de si se pondrá o no en servicio el terminal.

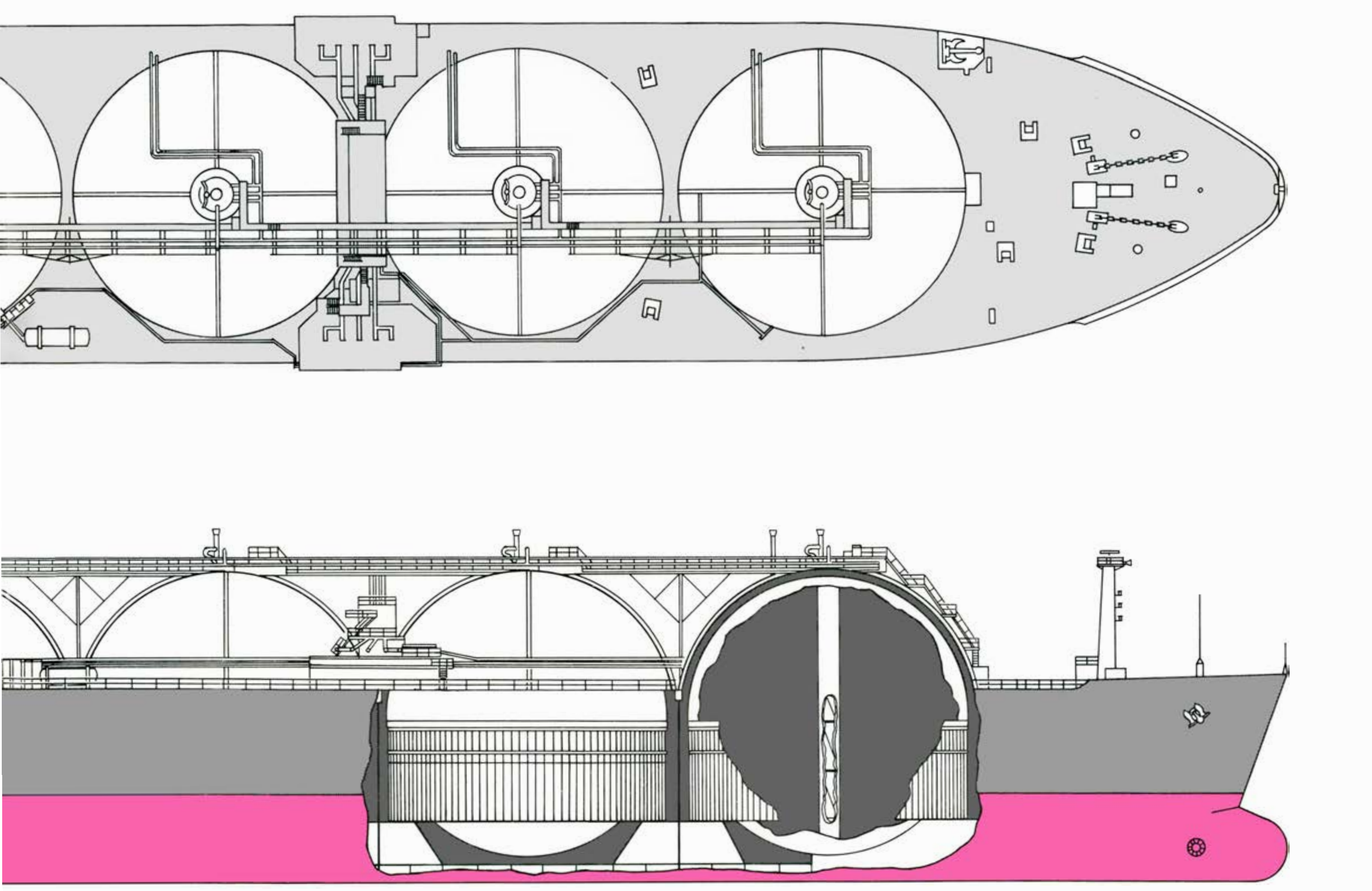
La mayor parte de los problemas relativos a la seguridad del gas natural licuado se refieren a los buques-cisternas y a los tanques de almacenamiento. Examinaremos algunos proyectos antes de volver al problema de la seguridad.

Los 35 buques-cisterna para gas natural licuado que están actualmente en servicio tienen una capacidad media de 35.000 metros cúbicos de líquido; otros 41 buques-cisterna que están en construcción o en la etapa de proyecto tienen una capacidad media de 124.000 metros cúbicos. Tal como hemos indicado, se están proyectando buques que alcanzarán una carga de hasta 165.000 metros cúbicos.

Los tanques de un buque que transporte gas natural licuado deben ser de doble pared, con una capa de aislamiento de más de un metro de espesor para disminuir la tasa de vaporización. Así pues, un choque que pudiera traer por consecuencia una pérdida de carga tendría que ser mucho más fuerte de lo que sería si el barco estuviese construido como un barco petrolero, en el cual el tanque es el propio casco del buque. Resulta muy improbable que ciertos accidentes, como las colisiones entre buques y las varaduras, provoquen la fuga de la carga de un buque que transporte gas natural licuado.

Los tanques de un buque destinado al transporte de gas natural licuado pueden construirse según uno de los tres proyectos básicos siguientes. El primero consiste en tanques autoportantes, separados del casco excepto en los elementos de sostén. El exterior de cada tanque está aislado, y puede dejarse una luz entre el aislamiento y el casco. El segundo diseño consiste en tanques de membrana, cuyo sostén procede del casco a través de aislamiento que soporta carga. En el tercer diseño se trata de tanques esféricos, construidos fuera del barco y que luego se hacen descender hasta encajar en unos espacios preparados durante las etapas finales de la construcción. En los tres tipos del tanque, el recipiente primario para el líquido está hecho de metal soldado. Además, el servicio de guardacostas de Estados Unidos exige que los tanques autoportantes y los de membrana posean una barrera criogénica secundaria. (Este requisito no rige para los tanques esféricos porque se les construye en condiciones en que puede comprobarse previamente la integridad de las soldaduras.)

En el mar, un buque-cisterna que transporte gas natural licuado presenta un perfil particular. La densidad de la carga no llega a la mitad de la densidad del agua del mar, por lo que el barco

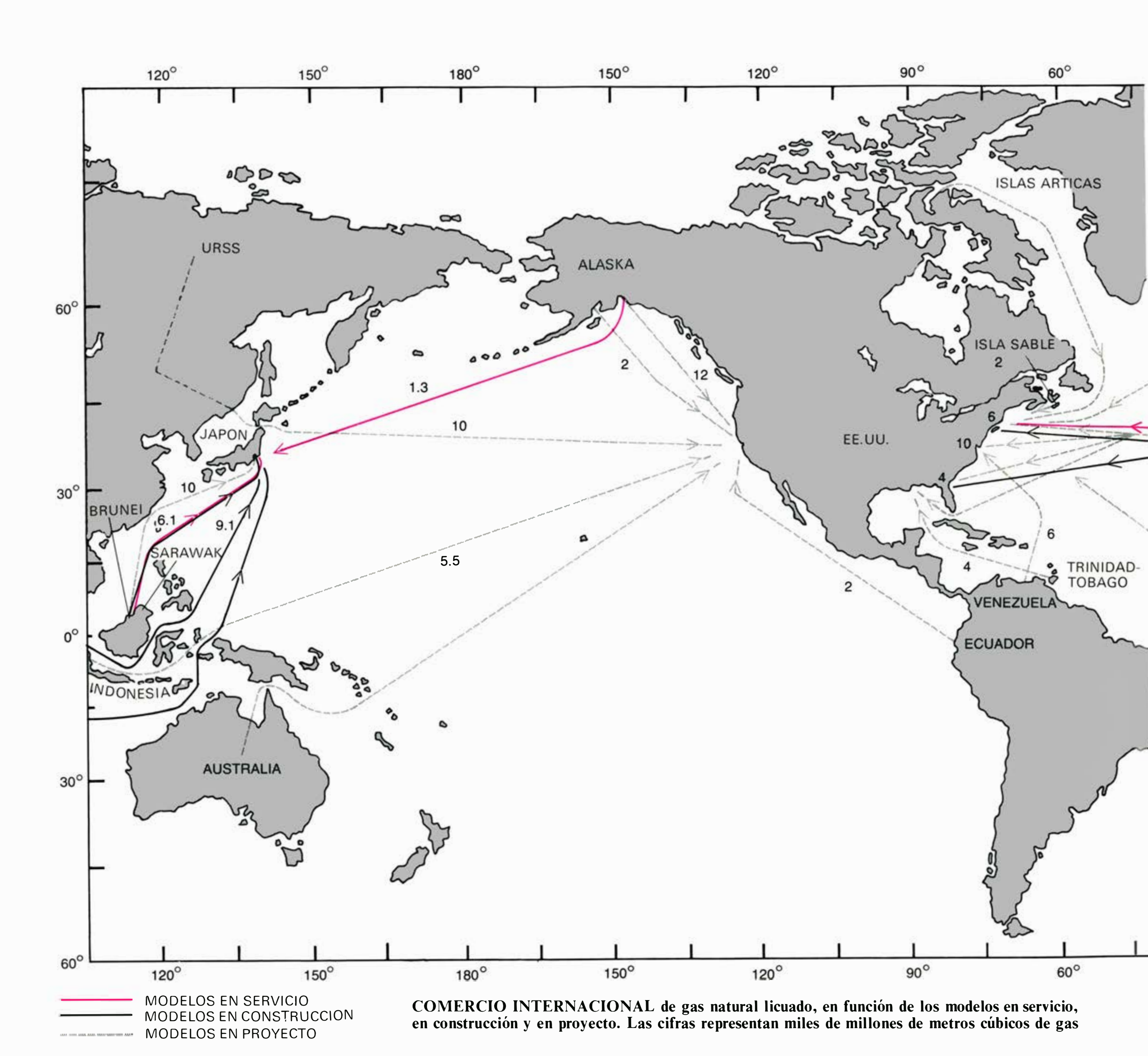


esféricos. Aparece en planta y de perfil. Cada tanque tiene un diámetro interior de 37,58 metros; está fuertemente aislado para minimizar la vaporización del frío líquido. El gas natural se deja vaporizar lentamente y se usa como combustible suplementario para las máquinas del buque.

navega muy poco hundido. Un barco típico tiene un calado de sólo 10 metros y su cubierta principal está entre 20 y 25 metros por encima del nivel del agua. Para proporcionar estabilidad en los viajes de retorno se necesitan tanques de lastre, que se rellenan de agua. Sin embargo, en esos momentos, el buque lleva un fondo residual de gas natural licuado, parte del cual se deja vaporizar lentamente para alimentar como combustible las máquinas del barco; el resto sirve para mantener baja la temperatura de los tanques, de modo que no sea preciso enfriarlos antes de cargar de nuevo. Puesto que construir y explotar un barco de tales dimensiones y características específicas es costoso, sus armadores tratan de minimizar sus períodos de inactividad. Esta consideración ha llevado a la creación de complejos y eficientes sistemas de carga y descarga. Unos tubos transportan gas natural licuado entre un barco y un tanque en tierra con altísimos caudales, hasta de cuatro metros cúbicos por segundo (50.000 galones por minuto).

Los tanques de almacenamiento deben estar hechos de materiales cuya resistencia se mantenga a –162 grados Celsius. El acero al carbono, material básico para la mayoría de las construcciones de acero, no resulta satisfactorio porque se vuelve frágil a baja temperatura. Pero se han realizado extensas pruebas, con gas natural licuado, en ciertos materiales, y se les ha encontrado adecuados para la construcción del tanque interior, que es el que de hecho retiene el helado líquido. Entre estos materiales se cuentan los aceros con gran proporción de níquel y algunas aleaciones de aluminio y hormigón pretensado. La capa aislante, que constituye un requisito para todo tanque en que se guarde gas natural licuado, se pone por fuera de la estructura que contiene el líquido y se mantiene dicho aislamiento en su lugar mediante un tanque exterior. El tanque exterior es lo que ve un observador. El aislamiento suele ser una capa de material incombustible llamado perlita.

Los tanques de almacenamiento para gas natural licuado construidos en un terminal para importaciones no son nunca pequeños. En un tanque típico, el diá-



COMERCIO INTERNACIONAL de gas natural licuado, en función de los modelos en servicio, en construcción y en proyecto. Las cifras representan miles de millones de metros cúbicos de gas

metro interior mide 55 metros, el diámetro exterior 58 metros y, 55 metros, la altura. Tiene capacidad para unos 90.000 metros cúbicos de gas natural licuado.

Cada tanque está rodeado de un dique de embalse, cuya capacidad de retención es, por lo menos, igual a la del tanque lleno. En muchos casos, el dique se ha construido a cierta distancia del tanque y, por tanto, es bastante bajo (de dos a diez metros de altura). En proyectos recientes se ha puesto el dique próximo al tanque. Uno de los motivos para hacerlo es que en caso de incendio, las llamas se confinarían en un área menor. Puede ser que un dique de esta especie sea tan alto como el tanque mismo. Además de exigir el dique en cuestión, las normas federales norteamericanas estipulan que, entre el dique y los límites de la instalación, debe dejarse una zona intermedia lo bastante ancha como para impedir otro desastre como el de Cleveland.

En las primeras etapas del desarrollo de instalaciones para gas licuado se pensó en almacenar el material en hoyos abiertos en el suelo y techados. Se preveía que el gas natural licuado crease una cavidad sellada al congelar el suelo adyacente. Pero la experiencia habida con tales tanques ha demostrado que se acumulan esfuerzos en el suelo congelado y se producen grietas. El incremento de la superficie de contacto entre el gas natural licuado y el suelo conduce a una mayor tasa de transferencia de calor hacia la cavidad, y el beneficio económico de la instalación llega a ser marginal. En Estados Unidos han sido abandonados cuatro tanques de este tipo; en otros países se siguen usando unos pocos, pero también se tiene previsto prescindir de su funcionamiento.

En Japón, en donde el terreno disponible es escaso y los terremotos constituyen un riesgo grave, se ha construido recientemente un gran número de tanques de hormigón soterrados para gas natural licuado. Por lo general, un tanque de este tipo lleva una capa aislante de espuma de polímero por dentro del hormigón. La capa de aislamiento sostiene una membrana que sirve como obturador o sello del gas licuado.

El gas natural licuado tiene una serie de interesantes características que provienen del hecho de que el metano, su constituyente principal, está mezclado con pequeñas cantidades de otros compuestos tales como etano, propano y nitrógeno. Es característica la "explosión de vapor sin llama" cuando entra en contacto con agua. David S. Burgess y sus colegas del Bureau of Mines, de Estados Unidos, fueron los primeros en dar noticia del fenómeno en 1970; lo observaron al medir la tasa de ebullición del gas natural licuado derramado sobre agua. En la prueba número 56 de una serie, que se había venido desarrollando sin incidentes, una explosión destruyó el tanque experimental que contenía el agua. Más adelante, una explosión parecida, pero de mayores proporciones, se produjo después de haberse derramado cerca de 250 decímetros cúbicos de gas natural licuado en un estanque abierto. En ninguno de ambos casos se incendió el material.

A fin de descubrir la causa de estas explosiones, Torr Euger y David E. Hartman, de la Shell Pipe Corporation, realizaron un amplio programa de ensayos. Llegaron a la conclusión de que las explosiones se debían a que una delgada capa de gas natural licuado, colocada sobre el agua, llegaba a calentarse de tal forma que la nucleación homogénea del vapor se producía con gran rapidez. Una explosión sin llama de vapor puede ocurrir sólo cuando el gas natural licuado lleva consigo fracciones significativas de etano y de propano. La gama de proporciones de composición es muy diferente de la del gas natural licuado típico, que conlleva solamente pequeñas proporciones de etano y propano.

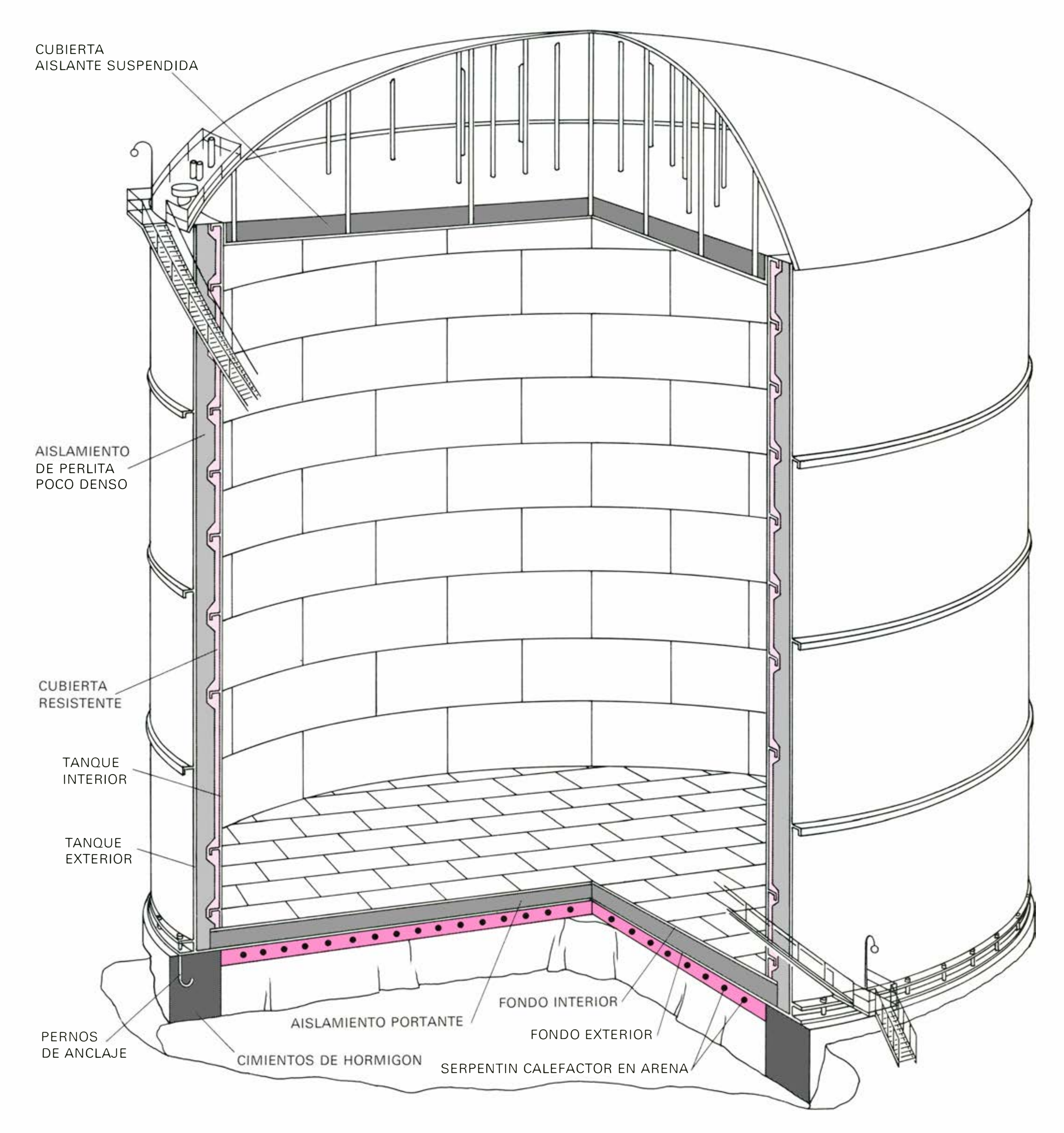
Aun cuando se produzca una explosión sin llama de vapor, el daño es leve. En el Liquid Natural Gas Research Center, del Instituto de Tecnología de Massachusetts, medimos las presiones existentes en el líquido próximo a la superficie en donde se produce la explosión

natural sin licuar. El gas natural licuado que se transporta o se proyecta transportar ocupa un volumen seiscientas veces menor que el ocupado por el gas natural cuando no se halla licuado.

y vimos que eran bastante inferiores a los 690 kilonewtons por metro cuadrado. Como la onda de presión se atenuaría rápidamente con la distancia, si se produjese una explosión de esa clase en el agua adyacente a un buque-cisterna tendría escaso efecto en el casco del barco. Las presiones en el aire dispuesto sobre la explosión son, por supuesto, menores.

Un segundo efecto interesante es el que las mezcla de gases naturales licuados de distinta composición pueden estratificarse. Este fenómeno se observó por primera vez en 1971, después de un incidente ocurrido en un terminal de importaciones en Italia. A las 18 horas después de que una carga de gas natural licuado hubiese sido puesta en un tanque de almacenamiento, la presión en éste empezó a aumentar. Pronto se pusieron en acción las válvulas de seguridad y, al cabo de una hora, se habían disipado al aire unos 150.000 kilogramos de gas natural. Afortunadamente, el gas evacuado no se incendió y la presión no subió tanto dentro del tanque como para ocasionar ningún daño grave.

El análisis del incidente reveló que la densidad de la nueva carga de gas na-



TANQUE DE ALMACENAMIENTO proyectado para contener gas natural licuado después que un buque-cisterna lo haya descargado en un terminal. Su capacidad es de 90.000 metros cúbicos. Está bien aislado, de manera que el gas natural licuado no se vaporice con excesiva rapidez. Los serpentines de calefacción que están dispuestos bajo el tanque impiden el alzamiento por helada, al variar la temperatura del suelo.

tural licuado era mayor que la del material que albergaba el tanque. Estaba asimismo más caliente. A causa de la diferencia de densidades, el nuevo material, que se había introducido en el fondo del tanque, permaneció allí, creando un sistema de dos capas. En las 18 horas subsiguientes, pasó calor al tanque a través de la base y de las paredes laterales, pues aunque el gas natural de la capa del fondo estaba más caliente que el de la capa superior, se encontraba de todas maneras mucho más frío que el entorno. Este tránsito de calor es normal, y en un tanque de almacenamiento no estratificado induce una circulación convectiva de gas natural licuado que mantiene bien mezclado, y a una temperatura casi igual, todo el contenido del tanque.

En un sistema de dos capas, el líquido de cada capa sigue circulando convectivamente, pero el fluido más denso no es lo bastante boyante como para penetrar en la capa superior menos densa. De este modo, puede acumularse energía en la capa más densa. Con el tiempo, la transferencia de calor y masa entre las dos capas tiende a igualar la diferencia de densidad. Entonces las dos capas pueden mezclarse rápidamente. Conforme el material trata de lograr el equilibrio térmico, la intrusión del material del fondo, más caliente, induce una rápida generación de vapor a fin de liberar la energía sobrante que se había acumulado mientras había dos capas. El contenido del tanque se mezcla rápidamente. Al fenómeno suele denominársele inversión.

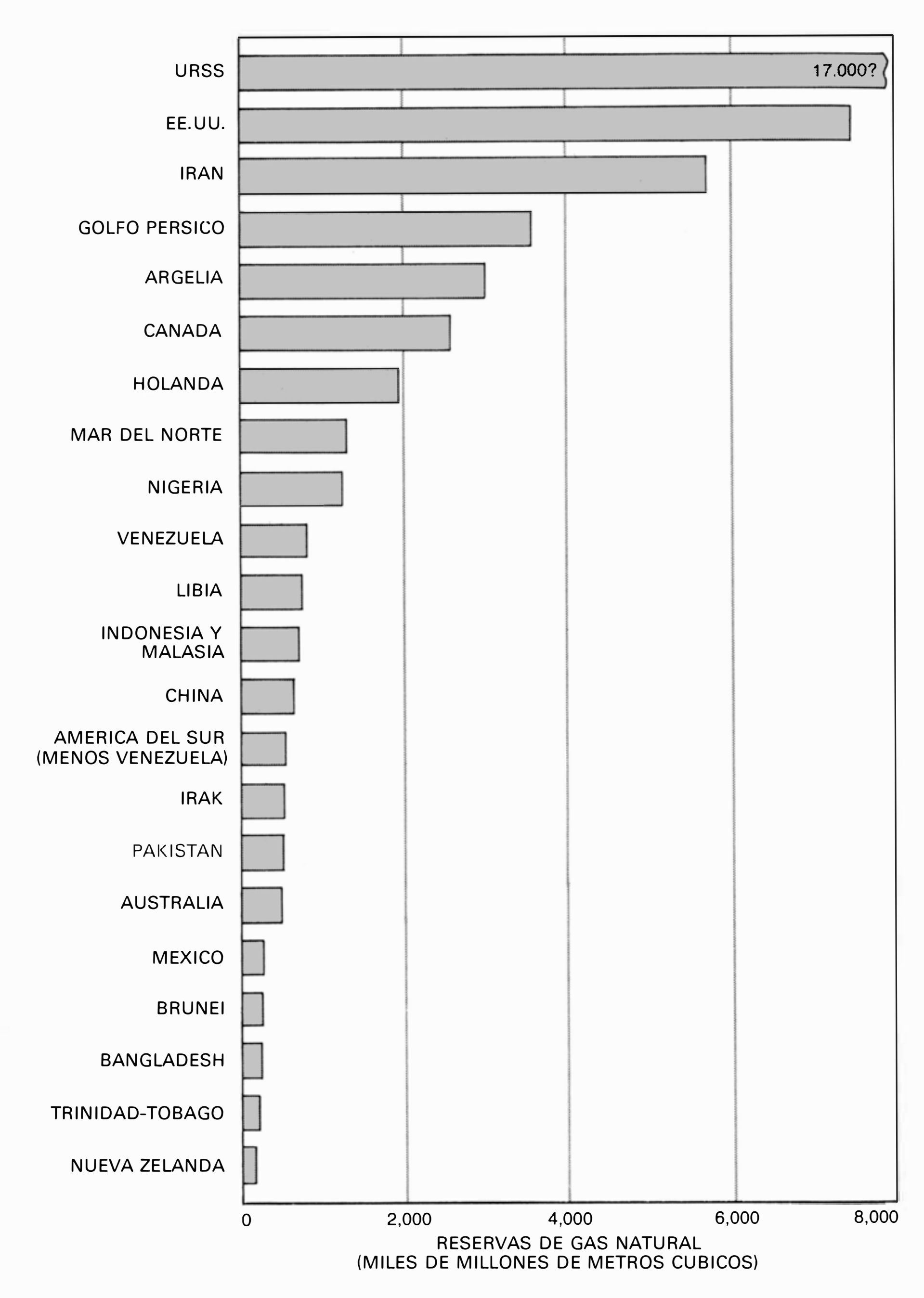
Las instalaciones modernas tienen procedimientos de norma para impedir la inversión. Un método consiste en echar el líquido menos denso al fondo del tanque y el más denso a la parte apical, con lo cual tiende a prevenirse la estratificación. Otra posibilidad es cargar siempre el tanque por el ápice; de este modo, si se produce estratificación, el tiempo necesario para la inversión es mucho más largo que el tiempo de permanencia del gas natural licuado en un tanque de almacenamiento de un terminal típico para importaciones de gas.

Se ha prestado gran atención a otros problemas, aparte de las explosiones sin llama de vapor y las inversiones. Si se derrama gas natural licuado, hierve rápidamente. Aunque el vapor no es tóxico, en una elevada concentración puede ocasionar asfixia al excluir el oxígeno. Además, la baja temperatura del material puede producir la congelación de tejidos orgánicos a cualquier persona que se halle en el campo de acción de una fuga. Estos dos posibles riesgos son demasiado localizados como para crear preocupación fuera de los límites de la instalación.

El principal riesgo del gas natural licuado es el fuego. Si se incendia una fuga, arde de modo muy parecido al incendio de un charco de gasolina. Si no se incendia pronto, el vapor inflamable puede ser transportado por el viento hasta encontrar una fuente de ignición. Experimentos realizados en Estados Unidos, Francia y Japón han mostrado que, una vez encendido el vapor, se forma un frente de llamas que retrocede ardiendo hacia la fuente de donde procede al vapor.

Otro asunto grave es el de si un incendio en una nube de gas natural licuado vaporizado podría convertirse en una detonación capaz de producir una onda explosiva perjudicial. El consenso entre los investigadores que han estudiado el problema es que son escasas las probabilidades de semejante explosión en un espacio no cerrado. No se ha confirmado la producción de una explosión al aire libre ni siquiera cuando se ha empleado como detonador una carga explosiva de gran potencia. Charles D. Lind, del Naval Weapons Center, no ha podido obtener detonaciones ni siquiera con cargas de hasta dos kilogramos.

Lind realizó sus pruebas con mezclas de metano y aire encerradas en hemis-



RECURSOS MUNDIALES de gas natural no licuado. Resulta improbable que gran parte de estas reservas se licúen y se exporten; muchos países, sin embargo, poseen un mercado interior pequeño o inexistente para el gas natural y, por tanto, tal vez consideren la posibilidad de exportarlo.

ferios muy finos de poliuretano de hasta 20 metros de diámetro. En el centro, y a nivel del suelo, colocó fuentes de ignición o cargas explosivas, siguiendo los resultados con fotografías ultrarrápidas. El recinto de poliuretano quedaba destruido, ya fuese por la detonación del explosivo de gran potencia o por la onda explosiva resultante, pero la onda de combustión del metano no se aceleró en ningún caso hasta convertirse en explosión. Se han demostrado explosiones de mezclas de aire y metano en recintos rígidos, pero normalmente la onda de combustión debe recorrer 30 metros o más antes de que pueda convertirse en una onda de detonación autosustentada.

El gas natural licuado encerrado en un tanque de almacenamiento es inocuo; no arde, a menos que se vaporice y se mezcle con aire. Unicamente en los escapes existe riesgo. Por eso mismo, los buques-cisterna y las instalaciones de almacenamiento están proyectados y funcionan de suerte que se prevengan los escapes fortuitos.

Los tanques para gas natural licuado están proyectados para cumplir con normas que especifican la intensidad del viento, terremotos, etc., que el tanque debe soportar. La norma suele ser la condición más grave experimentada durante los últimos 50 o 100 años. Los factores de seguridad, que también vienen establecidos en los reglamentos de construcción, garantizan que un tanque pueda soportar acontecimientos peores. Además, en una instalación de almacenamiento tiene que comprobarse que el público no se vería expuesto a ningún peligro si el tubo de mayor diámetro conectado al tanque se averiase mientras por su luz está pasando una corriente de gas licuado de caudal máximo.

Para un buque-cisterna de gas licuado que entra en un puerto se realizan parecidos análisis de los riesgos. Se recolectan datos sobre la intensidad del tráfico, y del tipo, velocidad y longitud de otros buques en aguas adyacentes. Como la probabilidad de un choque debería aproximarse a cero si se cumplen las normas de tráfico, los analistas suponen frecuentemente, y por prudencia, que todos los barcos se mueven de manera aleatoria. En cualquier caso, no todos los choques serían lo bastante graves como para resquebrajar el doble casco de un barco que transporte gas natural licuado hasta el punto de abrir un boquete por donde se fugase parte de la carga. Por tanto, el análisis se lleva más allá, a fin de calcular el momento mínimo y el ángulo de impacto necesarios para que un barco que choque ocasione una evacuación importante de gas natural ya licuado.

Solamente después de cuidadosos análisis se pueden calcular los posibles riesgos para el público de una operación de transporte de gas natural licuado. Por ejemplo, se calcula que el riesgo relacionado con la propuesta importación de gas natural licuado a Staten Island es de una fatalidad por cada 10 millones de años para las personas que viven o trabajan en la vía de entrada a la bahía. Ese nivel de riesgo es diez veces menor que el de morir por incendio en casa y aproximadamente igual al riesgo de ser alcanzado por un rayo.

Sin embargo, aunque puede proyectarse una instalación que proporcione un elevado grado de seguridad en caso del tipo de accidente contemplado en las normas de proyecto, cabe imaginar accidentes más graves provocados, por ejemplo, por un violento terremoto o por un minucioso acto de sabotaje. Podría preverse que la destrucción súbita de un tanque que contenga gas natural licuado diese por resultado un incendio importante, con unas consecuencias potencialmente desastrosas. ¿Puede tolerar la sociedad semejante riesgo?

El problema general no difiere de los problemas a que diariamente se enfrentan los ingenieros industriales y los organismos normativos del gobierno. Por ejemplo, ¿debería permitirse que un gran avión a reacción volase sobre una instalación pública multitudinaria, como un campo de fútbol? Existen unas posibilidades remotas, pero siempre finitas, de que se estrelle.

Ejemplos parecidos revelan que la sociedad tolera ciertos riesgos altamente improbables –riesgos que no pueden eliminarse con el proyecto– a fin de cosechar los beneficios de energía, bienes de consumo, transporte, sistemas centrales de abastecimiento de agua, etc. Fundándose en los estudios realizados acerca de los riesgos del gas natural licuado y en la experiencia que se ha adquirido en varios países sobre la manipulación de este combustible en gran escala, parece razonable afirmar que los riesgos son semejantes a los de la manipulación de otros combustibles como gasolina, propano (gas licuado de petróleo) y gas natural gaseoso. El gas natural licuado parece ser, por consiguiente, una prometedora fuente optativa de energía. Además, es de esperar que las nuevas experiencias que se adquieran sobre el tratamiento del gas natural licuado y las nuevas investigaciones realizadas por la industria y los organismos normativos estatales elevarán más el nivel de la seguridad en su uso.

La colisión entre la India y Eurasia

Durante los últimos 40 millones de años el subcontinente índico ha sufrido un empuje hacia el norte colisionando con la masa continental de Eurasia. Ello ha dado lugar a los terremotos más violentos y a las formas de relieve más variadas que se conocen

Peter Molnar y Paul Tapponnier

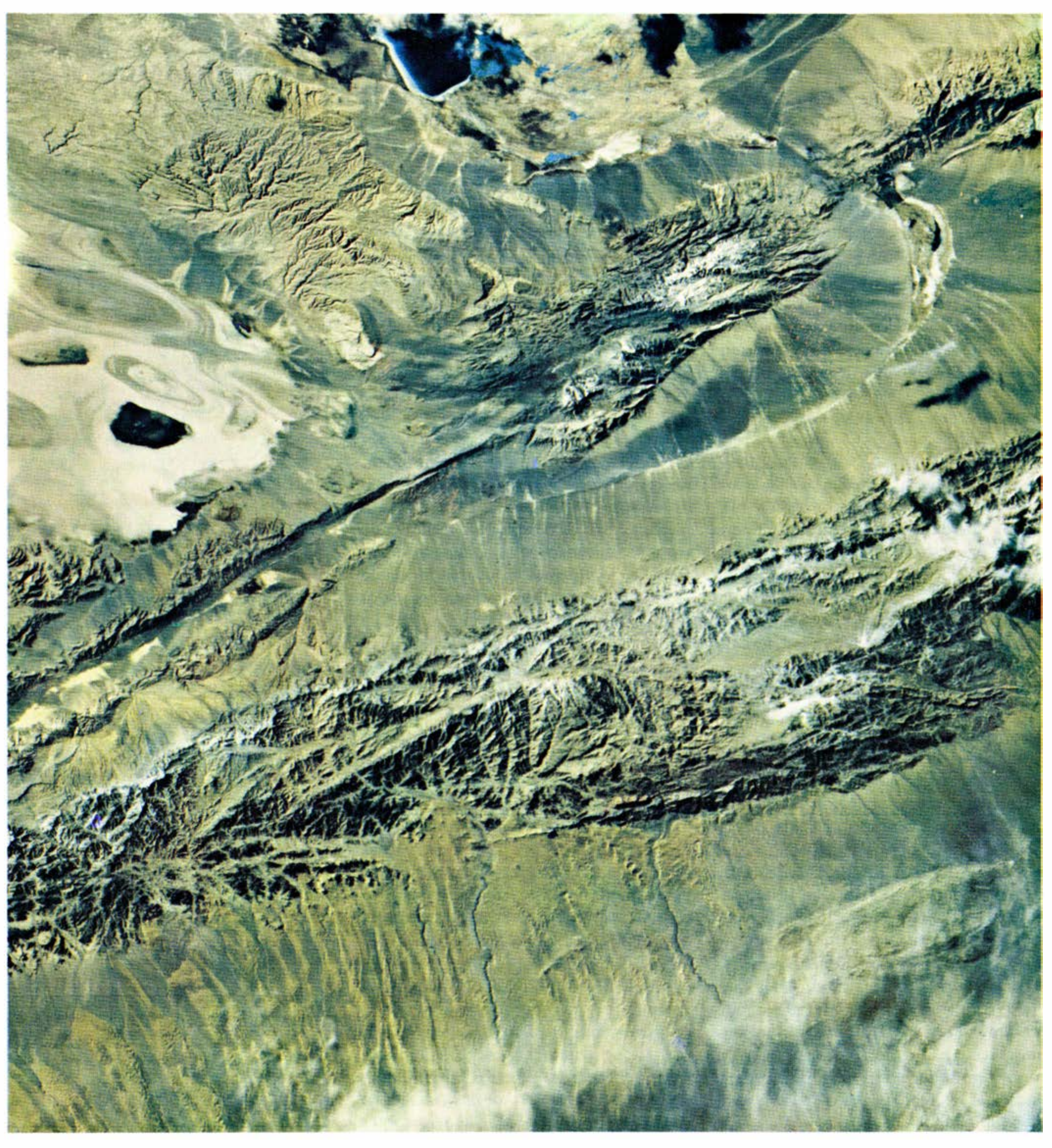
La región de la tierra que presenta la mayor diversidad geológica, topográfica y climática, así como una fuerte tendencia a sufrir movimientos sísmicos de primera magnitud, es la parte de Eurasia situada al este de los Urales y al norte del Río Ganges: comprende el norte de la India, Pakistán, Afganistán, el altiplano del Tíbet, Mongolia, la mayor parte de China y una buena porción de la parte oriental de la URSS. En esta región de características tan caóticas se yergue abruptamente la cordillera más elevada del mundo, el Himalaya, desde la llanura del Ganges, densamente poblada, al norte de la India, protegiendo el altiplano del Tíbet de la variación estacional de los vientos monzónicos que azotan el sur de Asia. El Tíbet tiene menos de 1,5 millones de habitantes repartidos sobre una superficie algo mayor que dos veces la de España y su altura media es de 5000 metros. En cambio, China oriental tiene abundantes precipitaciones y mantiene una población cercana a los mil millones de habitantes (se sabe que ya en 1556 un solo terremoto producido cerca de Sian, entonces la capital de China, causó 830.000 muertes). La amplia y elevada zona montañosa que abarca el Himalaya y el altiplano del Tíbet representa una barrera tal para las comunicaciones que, a lo largo de la historia, las poblaciones de la India y de China han tenido muy pocos contactos entre sí. Al norte del Tíbet, el desierto de Gobi constituye otra formidable barrera a la migración y a la comunicación. La cuenca de Tarim, que forma parte del desierto de Gobi, es una de las regiones más secas e inhóspitas de la tierra: en tres de sus costados, está rodeada por montañas de hasta 5000 metros de altura. Los vientos dominantes, que barren casi de continuo la cuenca, soplan con una fuerza tal que producen dunas de hasta 150 kilómetros de largo y cuyas longitudes de onda alcanzan de 3 a 5 kilómetros, claramente visibles en las fotos obtenidas por vía satélite. Al norte del desierto de Gobi se encuentra el lago Baikal, situado a 1800 metros de altura: se trata del lago más profundo de la tierra y ocupa la fosa del mismo nombre, que es una enorme grieta de la corteza terrestre, parecida a la existente en Africa oriental.

Si bien la geología de Asia parece presentar un barullo desordenado de formas de relieve, gran parte de la deformación superficial, cuando se la mira en conjunto con la ayuda de las fotos de satélite, parece ajustarse a un modelo coherente y simple, atribuible a una sola causa: la colisión geológica entre el subcontinente índico y el resto de Eurasia. Esta colisión continúa produciéndose hoy en día. Como ejemplo de sus efectos actuales creemos que el gran terremoto que devastó la ciudad industrial de Tang Shan el verano pasado fue producido por fuerzas que se originaban en la zona de colisión, a 2500 kilómetros al sudoeste.

A partir de los últimos diez años se ha analizado cuantitativamente la colisión. La hipótesis de la expansión de los fondos oceánicos se propuso, por primera vez, hacia la mitad de la década 1960-70 a fin de explicar el sistema de dorsales que forman la gran cordillera levantada sobre el fondo marino, que, con una longitud de unos 40.000 kilómetros, recorre las cuencas oceánicas del globo. Por la grieta que se produce en la cresta de la dorsal ascienden rocas fundidas procedentes del manto subyacente que rellenan el vacío creado en el fondo oceánico por el movimiento hacia fuera de las paredes laterales de la grieta. También se descubrió que el fondo oceánico está magnetizado en bandas de polaridad opuesta, correspondientes a la polaridad magnética que tenía la propia tierra en el momento en que la roca fundida cristalizó. (Por razones desconocidas aún, la polaridad magnética de la tierra se invierte a intervalos de varios centenares de miles de años.) El estudio de las inversiones magnéticas del fondo de los océanos Indico y Atlántico demostró que la India había recorrido 5000 kilómetros en relación a Eurasia antes de su colisión.

Desde el comienzo de la colisión, la India ha continuado moviéndose hacia el norte otros 2000 kilómetros con respecto a Eurasia. ¿Qué pasó con la enorme extensión de tierra que quedó desplazada por la colisión? Describiremos nuestra hipótesis para dar cuenta de ese desplazamiento y presentar las pruebas que la sustentan. La aceptación de la expansión de los fondos oceánicos, que confirmó las hipótesis anteriores acerca de la deriva continental, condujo con rapidez al concepto mucho más amplio de la tectónica de placas, el cual ha inspirado el renacimiento de las ciencias de la tierra. La tectónica de placas proporciona un mecanismo físico simple para explicar los movimientos horizontales a gran escala de las distintas partes de la corteza terrestre, permitiendo determinaciones cuantitativamente preci-

sas al describir la cinemática de la deriva continental. Uno de los conceptos centrales de esta teoría afirma que un número pequeño de grandes placas de la litósfera, la envoltura más externa de la tierra y de elevada resistencia, se mueven rígidamente una respecto de otra a velocidades comprendidas entre 1 y 20 centímetros por año, desplazándose sobre la astenósfera, que presenta una resistencia menor y una temperatura más elevada. Esta envoltura más externa tiene un espesor de 100 kilómetros aproximadamente; está formada por la corteza terrestre y la parte superior del manto. De ahí que, en tectónica de placas, a la corteza terrestre, más ligera y químicamente diferente del manto subyacente, se la vea como un elemento que forma parte de una placa litosférica y se mueve pasivamente con ella.



PARTE DE LA FALLA de Altyn Tagh, fotografiada desde 950 kilómetros de altura en febrero de 1973; está situada a 1200 kilómetros al norte del Monte Everest, en la región china de Sinkiang. Esta falla, que atraviesa la foto en diagonal desde el ángulo superior derecho hasta el inferior izquierdo, puede ser la mayor fractura del rumbo activa del mundo. La palabra "de rumbo" indica que los dos labios de la falla se han desplazado según sentidos opuestos entre sí. Es obvio que el labio meridional de la falla de Altyn Tagh se desplaza hacia la derecha, es decir, hacia el este, con respecto al labio septentrional, a consecuencia de la colisión de la India con el borde sur de Eurasia. La colisión empezó hace unos 40 millones de años y aún continúa. Aquí se observan 180 kilómetros de la fractura. La falla completa mide más de 2500 kilometros, si se incluye la falla de Kansu con la cual se une en su extremo oriental *(véase el mapa de las páginas* **22** y **23**)*.* **Esta fotografía y las de las páginas 25 y 27 se reproducen poniendo arriba el sur, de modo que los accidentes topográficos reciben la iluminación desde arriba. De otra suerte, las sombras apuntarían hacia la parte superior y el relieve tendería a invertirse; por consiguiente, los valles parecerían crestas.**

No deja de ser ironía que Alfred Wegener, probablemente el más importante entre los primeros defensores de la deriva continental, basara gran parte de su argumentación en la diferencia existente entre corteza continental y corteza oceánica. La corteza continental, que se mantiene muy por encima de la oceánica, se enraíza profundamente, pudiendo llegar hasta unos 35 kilómetros de la superficie mientras que, bajo los océanos, esta corteza tiene sólo unos 6 kilómetros de espesor. Wegener consideró a los continentes como poderosas naves navegando majestuosamente por la corteza y el manto, mucho más débiles, existentes bajo los océanos. Tal como se admite en tectónica de placas, lo que ocurre realmente es que la corteza junto con el manto superior (la litósfera) son extraordinariamente resistentes y, según parece, sólo se deforman en los bordes de placa; de acuerdo con lo cual, los movimientos relativos de las placas pueden describirse según los mismos postulados aplicables a los cuerpos rígidos.

Los continentes suelen comportarse como estructuras rígidas cuando se encuentran enteramente dentro de una placa. Sin embargo, si el límite entre dos placas pasa por el interior de un continente es, en general, más difuso que si se realiza entre dos placas oceánicas, o entre una placa oceánica y una placa continental, y, por tanto, es mucho más difícil de definir. Este carácter difuso se ve claramente en toda la zona mediterránea, siendo especialmente evidente en Asia, donde convergen las placas que forman la India y Eurasia.

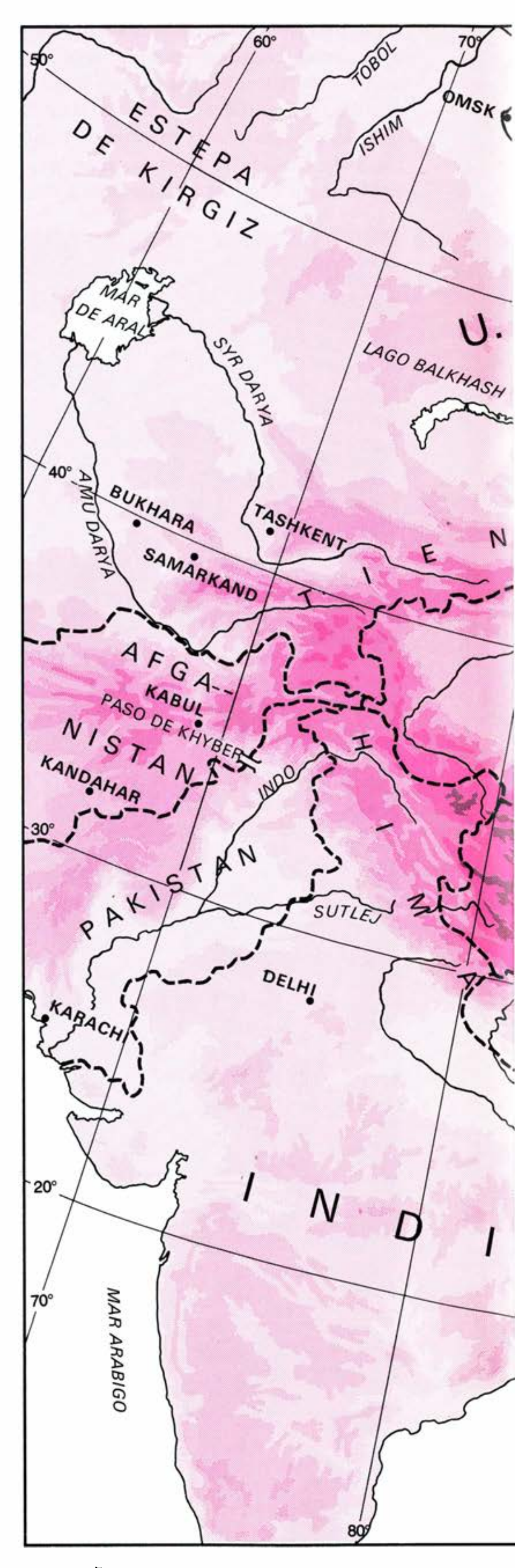
Antes de seguir adelante nos gustaría poner de relieve el importante papel que desempeña en tectónica de placas la diferencia existente entre continentes y océanos. La corteza oceánica se forma como parte de la litósfera en las dorsales oceánicas en expansión. La litósfera oceánica se enfría, se contrae y se hace más densa a medida que se aleja del centro de expansión. Una cantidad de corteza igual a la adicionada en la dorsal expansiva debe volver continuamente al manto. Así ocurre en las zonas de subducción, donde la litósfera oceánica se hunde en la astenósfera al ser empujada hacia abajo por ser más fría y, consiguientemente, ligeramente más densa que la astenósfera que la circunda. Si bien la densidad de la corteza oceánica es aproximadamente un 15 por ciento menor que la del manto, resulta lo suficientemente delgada como para ser arrastrada hacia abajo como parte de la litósfera. Tal es lo que acontece en muchos sitios, formando las típicas fosas oceánicas profundas, adyacentes a arcos de islas como los que bordean el Pacífico occidental.

Considérese, sin embargo, lo que ocurre cuando un continente suprayacente a una porción de litósfera descendente intenta hundirse en la astenósfera moviéndose a remolque de la litósfera oceánica consumida. La corteza continental, con unos 35 kilómetros de espesor y casi un 20 por ciento menos densa que el manto, tiene un grosor excesivo y resulta demasiado ligera para ser arrastrada hacia abajo, al interior de la astenósfera. Así pues, su capacidad de flotación impide que la corteza continental sufra una subducción, y explica tal vez el que, mientras la corteza oceánica experimenta un autorreciclaje cada 200 millones de años aproximadamente, hay extensas zonas continentales que han existido desde hace miles de millones de años, según es sabido.

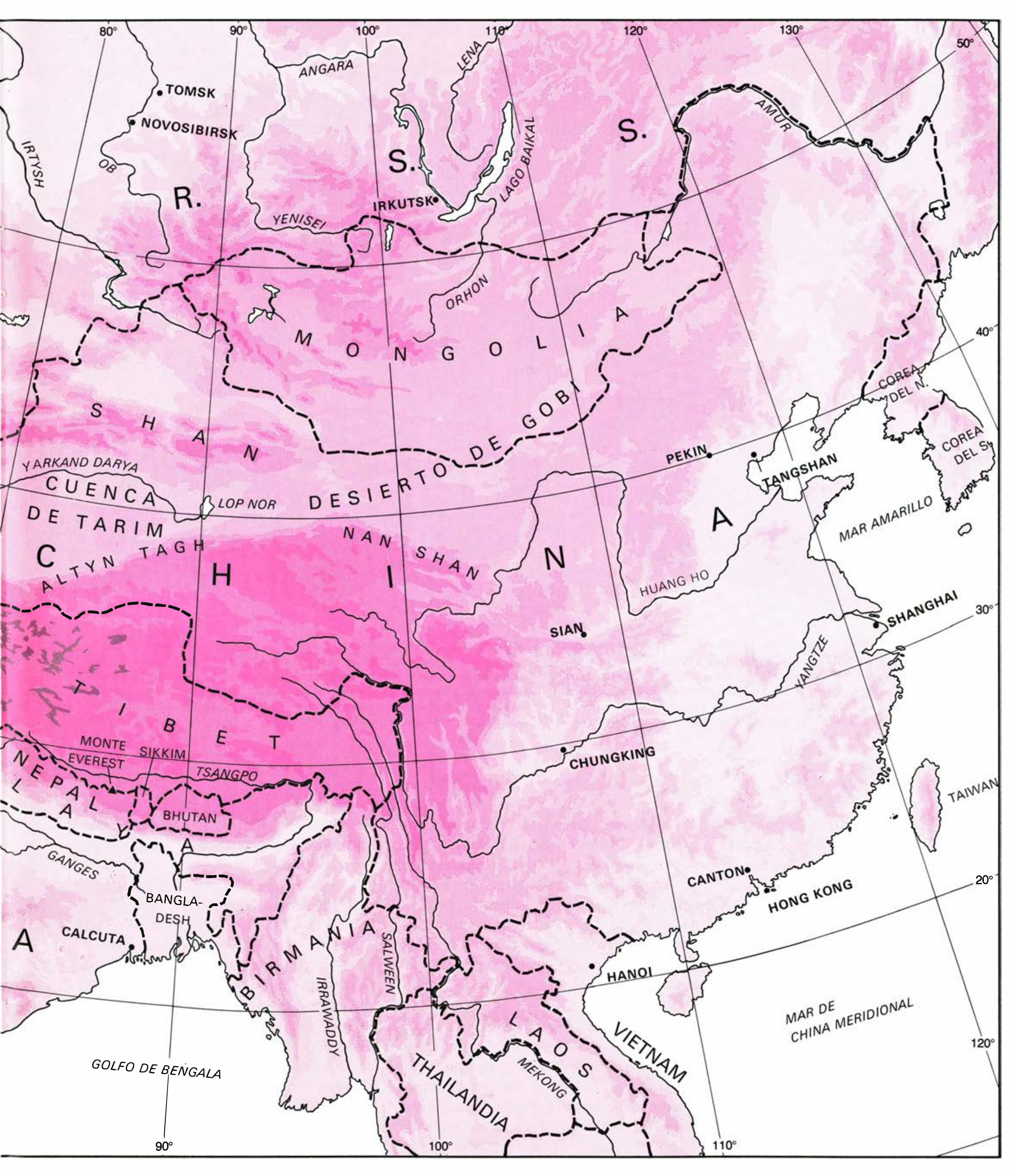
Cuando dos continentes colisionan, se sueldan para formar otro continente de mayores dimensiones. En ese caso, o bien cambia el movimiento relativo de las dos placas sobre las que se encuentran los continentes o bien se forma un nuevo límite entre ellas en un lugar distinto.

La historia geológica de la tierra nos proporciona varios ejemplos. Hace unos 250 millones de años, Europa y Siberia quedaron soldadas en los montes Urales y más o menos al mismo tiempo Norteamérica y Africa se encontraban unidas en los Apalaches meridionales antes de que se formara el actual océano Atlántico. Si observamos el mapa geológico de Asia se ven claramente varias suturas antiguas. Hace algunos años, Peter N. Kropotkin, el primer geólogo ruso de renombre que aceptó la deriva continental (cuyo tío-abuelo era el famoso anarquista del mismo nombre), describió Eurasia como un "continente compuesto", al que fueron uniéndose, sucesivamente, diversos fragmentos gracias a los fenómenos de colisión ocurridos durante los últimos 800 millones de años. La mayoría de las suturas de Asia tiene una antigüedad superior a los 200 millones de años, y sólo la existente entre la India y el resto de Eurasia es más moderna. Vamos a examinar esta reciente colisión y sus consecuencias.

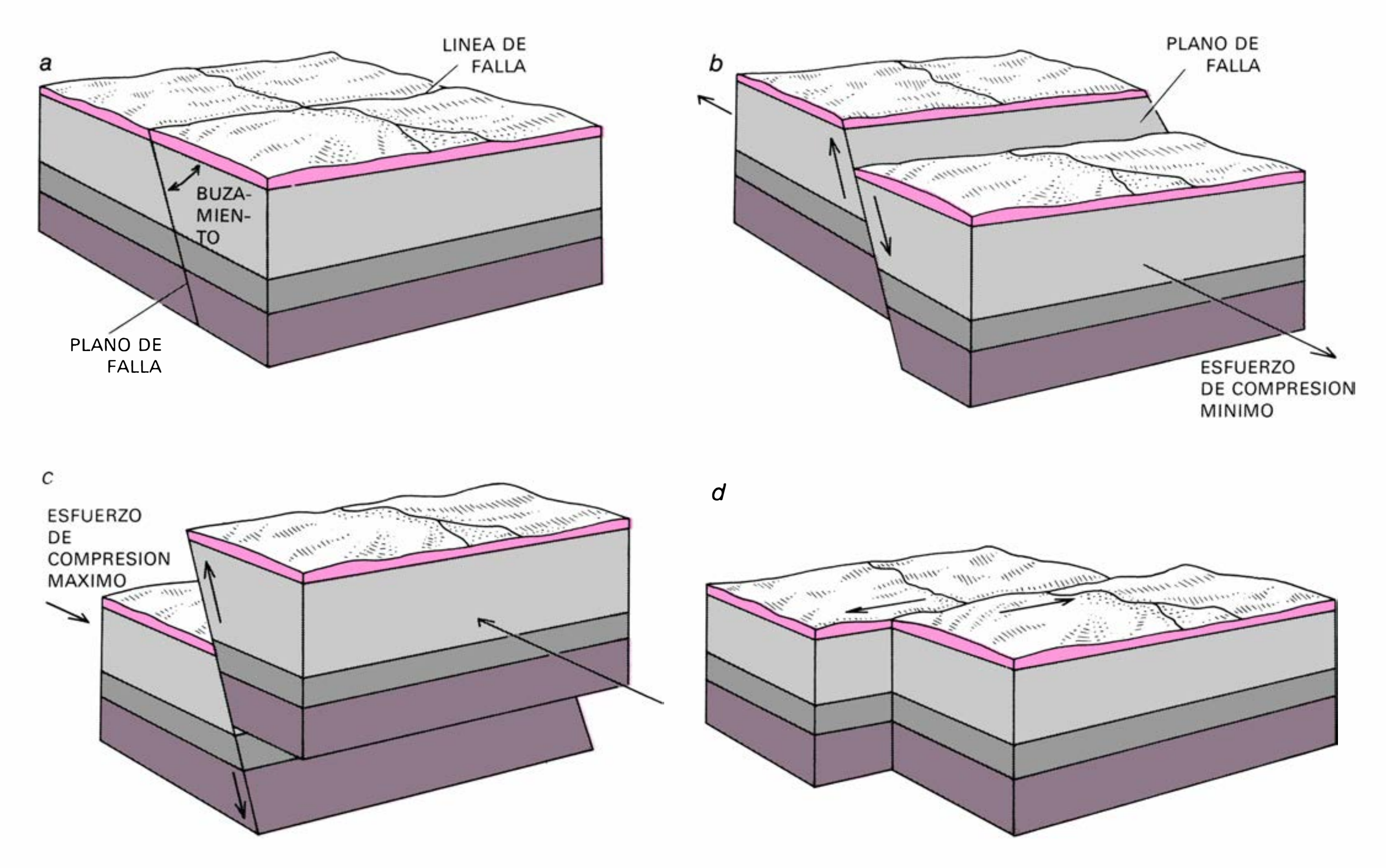
Aunque los procesos típicos producidos en tectónica de placas no parecen



LA COLISION entre la India y Eurasia ha comprimido y distorsionado la corteza terrestre desde el Himalaya a Siberia y desde Afganistán a la costa de China. Al norte del Himalaya, el Tíbet se ha elevado hasta una altura media de 5000 metros. El altiplano del Tíbet es drenado hacia el este por una serie de gran-



des ríos con valles casi paralelos al nacer, pero que más adelante se separan formando fértiles deltas extendidos según un arco que va desde la Bahía de Bengala hasta el Mar Amarillo. Si queremos cotejar la geografía y las distancias podemos imaginar un mapa de Norteamérica superpuesto sobre éste, de modo que el río Grande esté alineado con el Ganges y que nos colocaría Brownsville (Texas) cerca de Calcuta. (En realidad Calcuta está situada unos cuatro grados más hacia el sur que Brownsville.) Los Angeles coincidiría entonces aproximadamente con Kandahar (Afganistán), Miami estaría cerca de Hanoi y Portland (Maine) quedaría en las puertas de Pekín. El Himalaya trazaría un gran arco desde la frontera de Nevada-Utah hasta el Mississipi atravesando Arizona. Nuevo México, Texas y Louisiana. En el extremo norte, el lago Baikal cortaría el extremo más septentrional de Quebec. Mapa de Andrew Tomko.



DURANTE UN TERREMOTO, un bloque de la corteza terrestre se desliza respecto a otro adyacente siguiendo un plano de falla. En una falla normal *(b)* los bloques se comportan como si sufrieran una tracción. El bloque superior desciende siguiendo el buzamiento del plano de falla. En una falla inversa *(c)* ese labio superior es obligado a ascender siguiendo el buzamiento del plano de falla, ya que el esfuerzo de compresión máximo es horizontal y perpendicular. En una falla de rumbo *(d)* los dos bloques sufren un desplazamiento lateral respectivo. El desplazamiento lateral puede combinarse con movimientos de falla normal o inversa. A partir del análisis de las ondas sísmicas generadas por un terremoto, denominado solución en el plano de falla, se puede saber qué tipo de movimiento se ha producido en la fractura sujeta a estudio.

regir para la porción de Asia comprendida entre la India y Siberia, tectónicamente activa hoy en día, tal idea puede utilizarse para formular algunas restricciones sobre la historia del movimiento relativo registrado entre la placa de la India y la euroasiática. Podemos valernos de la constatación siguiente: si conocemos la historia del movimiento relativo entre la placa *a* y otras dos cualesquiera, *b* y *c*, podemos calcular la historia del movimiento relativo entre *b* y *c*. Gracias a las investigaciones geológicas llevadas a cabo en el Atlántico Norte por Jean Francheteau, del Centre Océanologique de Bretagne (Francia), y por Walter C. Pitman III y Manik Talwani, del Lamont-Doherty Geological Observatory, de la Universidad de Columbia, sabemos cómo se han movido Eurasia y Africa respecto de Norteamérica, y podemos, por tanto, calcular cómo se movieron una respecto de la otra. Análogamente, los trabajos de Robert L. Fischer, de la Scripps Institution of Oceanography, D. P. McKenzie, de la Universidad de Cambridge, y John G. Sclater, del Instituto de Tecnología de Massachusetts, han permitido conocer las características del movimiento relativo entre la India y Africa; por tanto, estamos en condiciones de calcular la posición de la placa índica con respecto a Eurasia en distintos momentos del pasado.

Se desconoce la posición del borde norte del continente índico original con respecto a la India actual, debido ello a que este margen se ha visto intensamente deformado por la formación del Himalaya. Sin embargo, si se trata de calcular el valor del movimiento relativo entre la placa de la India y la de Eurasia a diferentes intervalos de tiempo, el desconocimiento de ese detalle carece de importancia. Además, aunque desconozcamos la distancia entre este margen antiguo y el norte de India, a partir de la geología de Asia podemos determinar la posición del límite entre las rocas del subcontinente índico actual y las que formaban parte de Eurasia mucho antes de la colisión.

La primera prueba utilizada para delimitar la sutura entre dos continentes es la existencia de una secuencia de rocas conocidas con el nombre de serie ofiolítica. Las ofiolitas poseen tres características particulares. Presentan una secuencia sedimentaria característica en la que se incorporan rocas estratificadas silíceas, tipo "chert", que son peculiares de la sedimentación oceánica profunda. Contienen, además, restos de basaltos almohadillados, que son rocas ígneas de formas apelmazadas, típicas de las lavas basálticas submarinas como las emitidas en las dorsales oceánicas en expansión. También se incorporan a la secuencia rocas densas, oscuras, con bajo contenido en sílice, conocidas con el nombre de rocas ultramáficas, consideradas como típicas del manto. Las ofiolitas se interpretan como escamas formadas por corteza oceánica y manto superior; de ahí que su

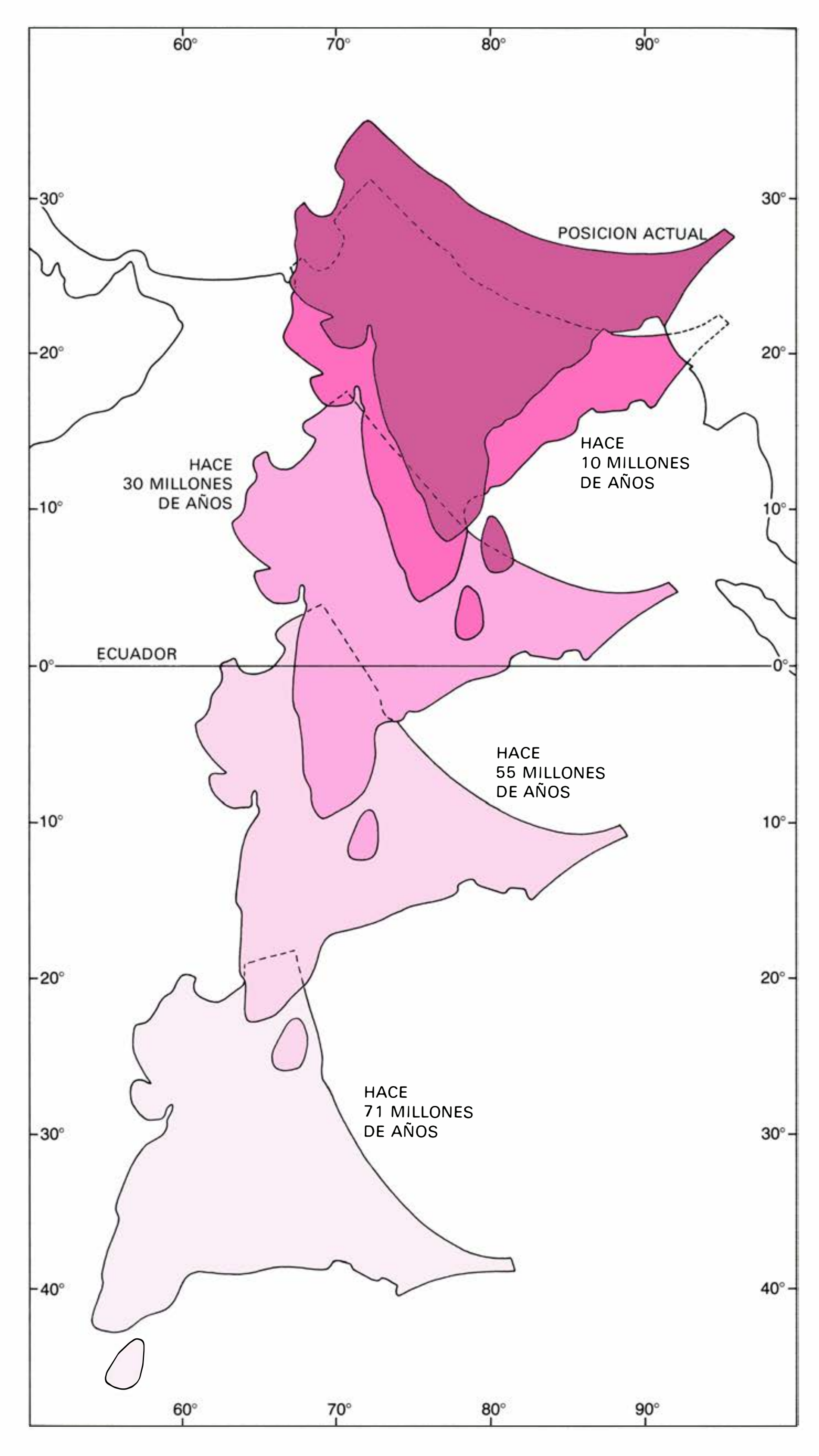
presencia implique la existencia anterior de una cuenca oceánica.

Una banda de ofiolitas, que sigue el curso de los valles de los ríos Indo y Tsangpo, en el Tíbet meridional, al norte del Himalaya, señala el contacto entre los continentes soldados. Algo más hacia el norte existen rocas volcánicas típicas de las zonas de subducción, tal como ocurre en los Andes americanos. De la misma forma que hoy en día la corteza oceánica se hunde hacia el este bajo los Andes, resulta muy posible que el fondo oceánico antiguo existente entre la India y Eurasia se hundiera hacia el norte por debajo del Tíbet. En cambio, al sur de la sutura, la cordillera del Himalaya es el fruto de un apilamiento de escamas, constituidas por los materiales que formaban la antigua parte septentrional de la India. Patrick Le Fort, del Centre de Recherches Pétrographiques et Géochimiques (Francia), y Maurice Mattauer de la Universidad de Montpellier han demostrado que hay una progresión de norte a sur en el apilamiento de esas escamas, de modo que el accidente cabalgante más antiguo está situado al norte.

La próxima etapa consistirá, probablemente, en la formación de una nueva falla más meridional, en la llanura del Ganges, y en el corrimiento, hacia el sur y sobre esta última, de los materiales que se encuentran al norte.

Aunque la geología del Himalaya y del Tíbet no es suficientemente bien conocida como para recomponer la posición del borde septentrional del continente índico original, lo que sí nos permite establecer es una serie de restricciones importantes acerca del problema de la época en que posiblemente colisionaron la India y Eurasia. Existen cuatro tipos de observaciones que implican una edad comprendida entre los 40 y los 60 millones de años. En primer lugar, gracias a las investigaciones geológicas de las ofiolitas de esta zona, realizadas por Augusto Gansser, del Instituto Federal Suizo de Tecnología, se encontraron "bloques exóticos" del Cretácico superior (hace unos 70 millones de años) incluidos en la serie. Así pues, las ofiolitas han tenido que formar parte de un fondo oceánico hasta una época de datación más tardía.

En segundo lugar, Gansser describe, en el borde norte del Himalaya, una secuencia de rocas sedimentarias, típica de la plataforma y talud continentales que empieza en el Cámbrico (hace unos 500 millones de años) y se continúa hasta el Eoceno inferior (hace unos 50



LA DERIVA DE LA INDIA HACIA EL NORTE ha sido reconstituida a partir de las inversiones magnéticas de los fondos oceánicos del Indico y Atlántico. A medida que la roca fundida ascendía hasta la dorsal existente en el fondo oceánico y se solidificaba, se magnetizaba de acuerdo con la polaridad que el campo magnético terrestre tenía en aquel momento. La polaridad de la tierra cambia a intervalos esporádicos e irregulares, dejando un registro que puede datarse. Esta reconstrucción a base de "lapsos de tiempo" demuestra que la India se desplazó unos 5000 kilómetros hacia el norte con respecto a Eurasia durante los 20-30 millones de años anteriores a su colisión con ésta. Durante los últimos 70 millones de años el extremo nororiental ha sufrido un desplazamiento efectivo de unos 7000 kilómetros. La velocidad del continente en conjunto fue de unos 10 centímetros por año durante los primeros treinta millones de años y de 5 para los siguientes 40 millones. La reconstrucción supone que el borde de Eurasia se mantuvo en su posición actual.

millones de años). Así pues, resulta que ha existido un océano entre los continentes convergentes hasta aproximadamente esta última fecha.

En tercer lugar, no se conocen en la India mamíferos fósiles cuya edad sea superior a los 50 millones de años. Ashok Sahni y Vimal Kumar, de la Universidad de Lucknow (India), indican que los mamíferos más antiguos encontrados en ese país pertenecen al Eoceno medio (hace unos 45 millones de años) y se parecen a los hallados en Mongolia. Por consiguiente, aunque los mamíferos hubieran evolucionado en otros continentes, parece que no lo hicieron de forma independiente en la India, con lo que se pone de relieve que el continente permaneció aislado hasta hace unos 45 millones de años.

La colisión permitió, pues, que invadieran la India una gran cantidad de mamíferos procedentes de Mongolia. En cuarto lugar, según Gansser la fase principal de formación del Himalaya empezó en el Oligoceno.

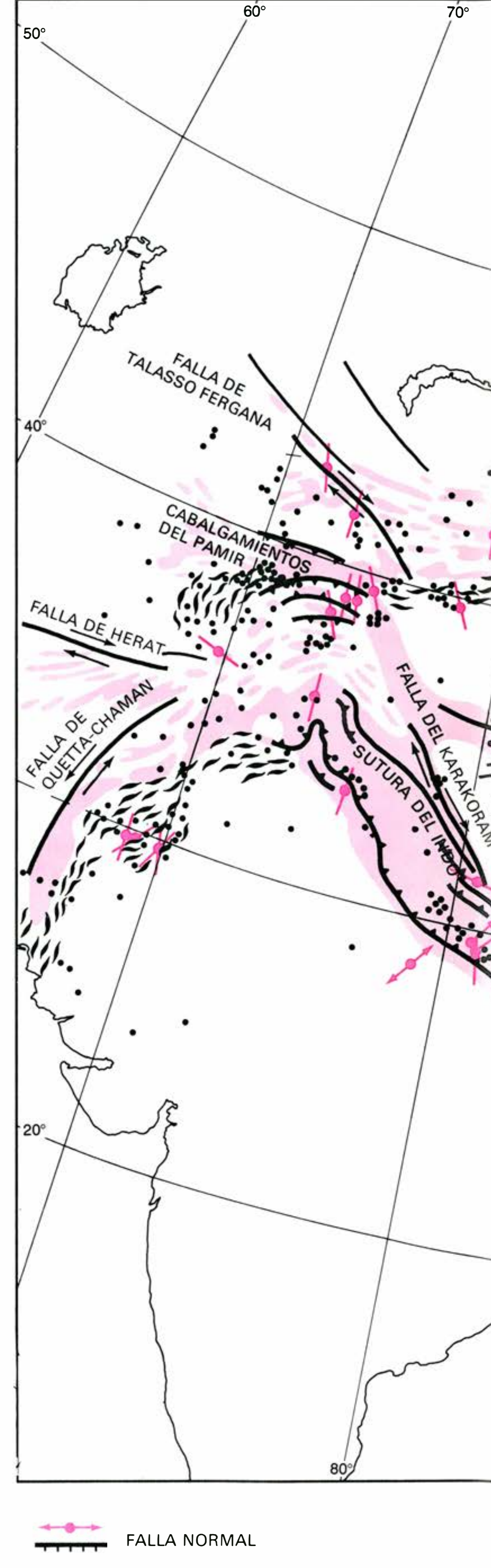
Estas observaciones no nos permiten fechar con precisión la época exacta de colisión entre la India y Eurasia, pero pensamos que aconteció probablemente hace 45 millones de años (con unos cuantos millones de más o de menos). Sin embargo, es muy poco probable que los antiguos márgenes de la India y Eurasia entraran en contacto simultáneamente en toda su longitud. Es más probable que su primer contacto se produjese entre penínsulas y que en el transcurso del tiempo la zona de contacto aumentara hasta hacer desaparecer la cuenca oceánica existente entre los continentes. Consideramos, pues, probable que el contacto inicial haya podido ocurrir unos pocos millones de años antes que el contacto completo.

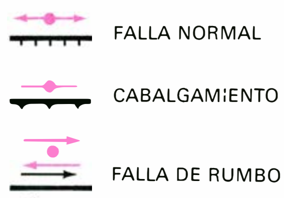
A partir de la historia de la expansión del fondo del Océano Indico se pueden calcular las posiciones relativas de la India y Eurasia a lo largo de algunas decenas de millones de años. Puede verse inmediatamente que entre hace 70 y 40 millones de años la India recorrió, con respecto a Eurasia, unas dos veces la distancia ganada a partir de entonces. Un gráfico de la distancia de los extremos nororiental y noroccidental de la India en diferentes épocas del pasado respecto de sus posiciones actuales muestra que la velocidad de convergencia entre estas dos partes varió según un factor de dos, hace unos 40 millones de años. Debido a la incertidumbre de los datos disponibles, el cambio de velocidad pudo haberse producido unos pocos millones de años más tarde o bien incluso 10 millones de años antes. En todo caso, interpretamos el cambio en la velocidad de acercamiento a la posición actual de la India como señal de que las primeras etapas de la colisión se produjeron más o menos simultáneamente con ese cambio de velocidad y de que la corteza continental de este subcontinente, en vez de sufrir una subducción, frenó el movimiento de la placa hacia el norte.

Si bien las reconstrucciones respaldan la teoría de que la corteza continental no puede sufrir subducción, nos plantean otro problema posiblemente más difícil de resolver. Si aceptamos que la India y Eurasia colisionaron hace 40 millones de años, debemos aceptar igualmente que, desde esa época, la India se ha ido desplazando hacia el norte, con respecto a Eurasia, unos 2000 kilómetros. Si los continentes colisionaron antes, la distancia recorrida es incluso mayor. Teniendo en cuenta que la corteza continental no puede sufrir subducción nos enfrentamos con el problema de explicar el desplazamiento de una parte de la corteza cuya extensión tiene la anchura de la India y 2000 kilómetros de longitud.

El movimiento continuado de la India hacia el norte, a una velocidad de unos cinco centímetros por año, es, probablemente, el responsable de la actividad tectónica generalizada existente en Asia. Por ejemplo, la actividad sísmica se hace sentir en toda una zona que llega hasta unos 3000 kilómetros al norte y al este del Himalaya. Entre los 22 terremotos de intensidad máxima del inventario que Beno Gutenberg y Charles F. Richter, del Instituto de Tecnología de California, realizaron para el período 1897-1955, siete se produjeron en Asia central y oriental de los cuales cuatro ocurrieron al norte del Himalaya. La deformación de la superficie terrestre que acompañó a algunos de estos terremotos fue enorme. El registrado en 1957 en la región de Gobi-Altai (Mongolia), después de haberse publicado el trabajo de Gutenberg y Richter (que tampoco, seguramente, se hubiera incluido en él por ser demasiado débil), produjo desplazamientos de hasta 10 metros a lo largo de la falla principal asociada al movimiento sísmico. Estos graves desplazamientos a lo largo de fallas parecen constituir una característica de varios de los grandes terremotos ocurridos en Asia. Sea lo que fuere, resulta evidente que la región, considerada globalmente, no se comporta como una placa rígida.

En Asia, además de las elevadas montañas que forman el Tíbet y el Himalaya,





80° 90° 100° 110° 130°

40° 30° 20°

90° 100° 110°

SISTEMA DE FOSAS DEL BAIKAL

FALLA DE DZUNGARIAN

FALLA DE BOLNAI

ALTAI

TIEN SHAN

FALLA DE BOGDO

CUENCA DE TARIM

SISTEMA DE FOSAS DE SHANSI

FALLA DE ALTYN TAGH

FALLA DE KANSU

NAN SHAN

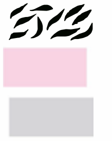
FALLA DE KUNLUN

CABALGAMIENTO DE LUNG MENG SHAN

FALLA DE KANG TING

CABALGAMIENTO FRONTAL HIMALAYO

FALLA DEL RIO ROJO



PLIEGUES

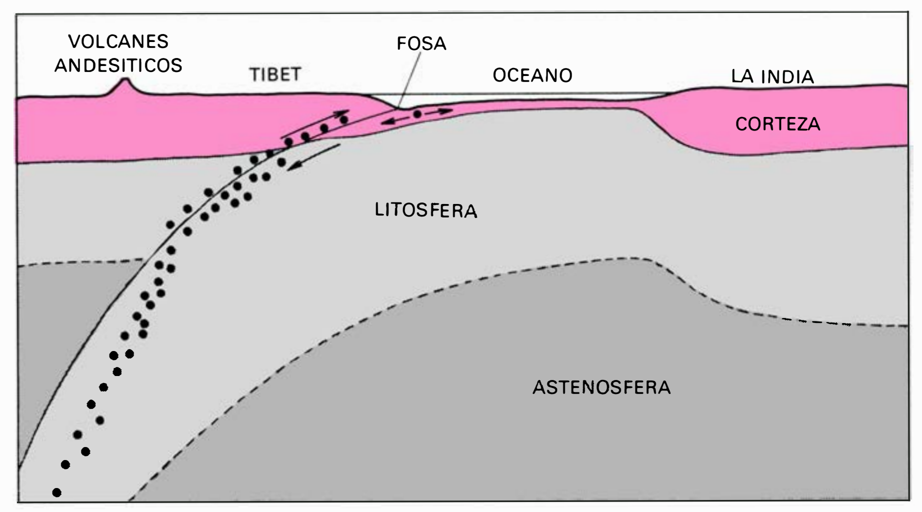
ZONAS EN COMPRESION

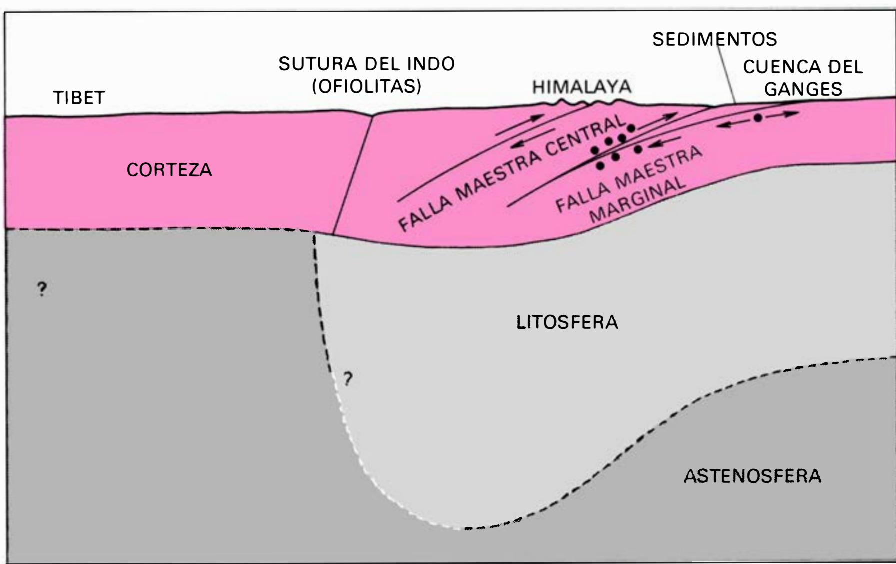
ZONAS EN EXTENSION

TERREMOTOS MENORES

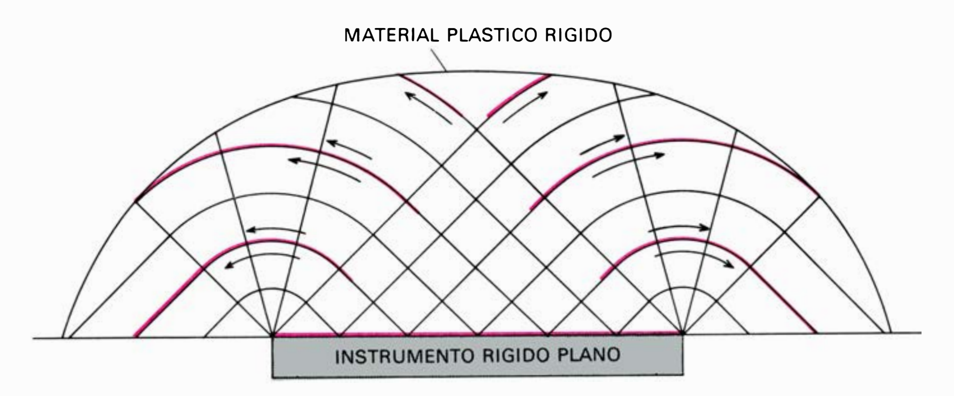
GRANDES TERREMOTOS

LOS PRINCIPALES RASGOS TECTONICOS que se consideran asociados al empuje continuado hacia el norte de la placa índica contra la placa de Eurasia han sido representados por los autores basándose en parte en el análisis de las fotos del ERTS y en parte en estudios de los principales terremotos *(puntos coloreados),* los cuales revelan cómo se ha movido la corteza a lo largo de las fallas. Las rectas sin flecha que atraviesan puntos indican fallas inversas, las flechas dobles, fallas normales y los pares de flechas antiparalelas el movimiento según fallas de rumbo. Las zonas en color son zonas de levantamiento reciente, producido por un acortamiento cortical. La impresión global es que la gran masa continental de Eurasia situada al oeste del meridiano de longitud 70° este ha permanecido más o menos sin deformar, conforme China era empujada hacia oriente.





CORTE TRANSVERSAL DE LA COLISION entre las placas índica y euroasiática. El diagrama superior muestra un corte a través de la litósfera y la astenósfera, unos dos millones de años antes de producirse el contacto, cuando las masas continentales estaban separadas aún por unos 200 kilómetros de océano. En esta época, la placa litosférica portadora de la India se hundía por debajo de la placa eurasiática del mismo modo que la placa del Pacífico se hunde hoy en día por debajo de la placa sudamericana, originando a los Andes, a lo largo de la costa oeste de América del Sur. Los puntos negros muestran la tendencia de los terremotos a agruparse a lo largo del contacto entre placas y en el interior de la placa descendente. El diagrama inferior muestra la situación actual. La línea de sutura, denominada del Indo, se caracteriza por la presencia de ofiolitas, que son secuencias de rocas con sedimentos océanicos y otros caracteres que demuestra su formación en un medio suboceánico. Los terremotos se distribuyen de forma más difusa; son más superficiales que antes de la colisión. El Himalaya está formado por escamas de antigua corteza índica que han cabalgado sobre el resto de la India situado al sur, dando lugar a nuevas fallas que migran hacia la parte meridional. La falla maestra original parece ser activa. Bajo el Tíbet, la corteza se muestra muy caliente; quizá, la litosfera se haga allí tan delgada que su base se encuentre en la corteza.



GEOMETRIA DE LAS LINEAS DE DESLIZAMIENTO observadas cuando un objeto hecho de un material duro como el acero, se comprime contra un material más blando como el latón. Este símil tiene un notable parecido con la distribución y sentido direccional de las fallas de rumbo de Asia. Las líneas de deslizamiento son probablemente menos simétricas en Asia que en el caso ideal.

existen otras, más al norte y al este, como las cordilleras de Tien Shan y Nan Shan, que tienen picos de hasta 6000 metros. Las montañas de gran altura suelen desgastarse rápidamente en virtud de la erosión, de suerte que su existencia implica grandes movimientos corticales producidos en tiempos geológicos recientes. Los estudios abordados por geólogos rusos (V. N. Krestnikov, A. V. Goryachev, S. A. Zakharov y otros) demuestran que la zona de la cordillera de Tien Shan era casi llana, entre los 200 y 30 o 40 millones de años antes de la actualidad y que se ha ido elevando progresivamente desde entonces. Si bien otros trabajos consultados no son tan concluyentes, lo que sí sugieren es que el relieve actual de Mongolia y China es relativamente reciente. Por tanto, la deformación de una gran parte de la corteza de Asia parece haber empezado después de su colisión con la India. Podemos deducir que la penetración de esta última en Eurasia fue responsable de la deformación.

Si intentamos explicar cómo pudo la India desplazarse 2000 kilómetros, después del inicio de su colisión con Eurasia, dispondremos de un grado de libertad mayor para hacerlo si imaginamos que el material trasladado se absorbió en un área de varios millones de kilómetros cuadrados y no en una zona restringida, cercana a la sutura. La formación del Himalaya puede justificar solamente una parte del material desplazado; queda, pues, por explicar dónde ha ido a parar el resto. Un método que ha demostrado ser particularmente útil para decidir cómo se desplaza el material en las regiones oceánicas actuales es la determinación de las denominadas "soluciones en el plano de falla" de los terremotos. Estas soluciones ponen de manifiesto la dirección del movimiento relativo a lo largo de la falla. A partir del estudio de las ondas emitidas por un terremoto se puede determinar el tipo de falla que se produjo durante el movimiento sísmico y la dirección de movimiento de un labio de la fractura con respecto al otro.

Los geólogos clasifican las fallas en tres clases principales. Se produce una falla normal cuando los esfuerzos en posición horizontal producen el hundimiento de un labio de la fractura en relación al otro. Una falla inversa se produce cuando los esfuerzos compresivos hacen montar uno de los labios de la falla sobre el otro. En una falla de rumbo los dos labios sufren un desplazamiento horizontal respectivo. Si para un observador situado en cualquiera de los labios el otro

se ha movido hacia la izquierda, se dice que el desplazamiento ha sido lateral izquierdo (o levógiro). Si el movimiento ha sido en sentido opuesto, el desplazamiento es lateral derecho (o dextrógiro). En colaboración con Thomas J. Fitch, actualmente en el Lincoln Laboratory del Instituto Tecnológico de Massachusetts, y Francis T. Wu, de la Universidad estatal de Nueva York en Binghamton, hemos recopilado las soluciones en el plano de falla de 75 terremotos ocurridos en Asia.

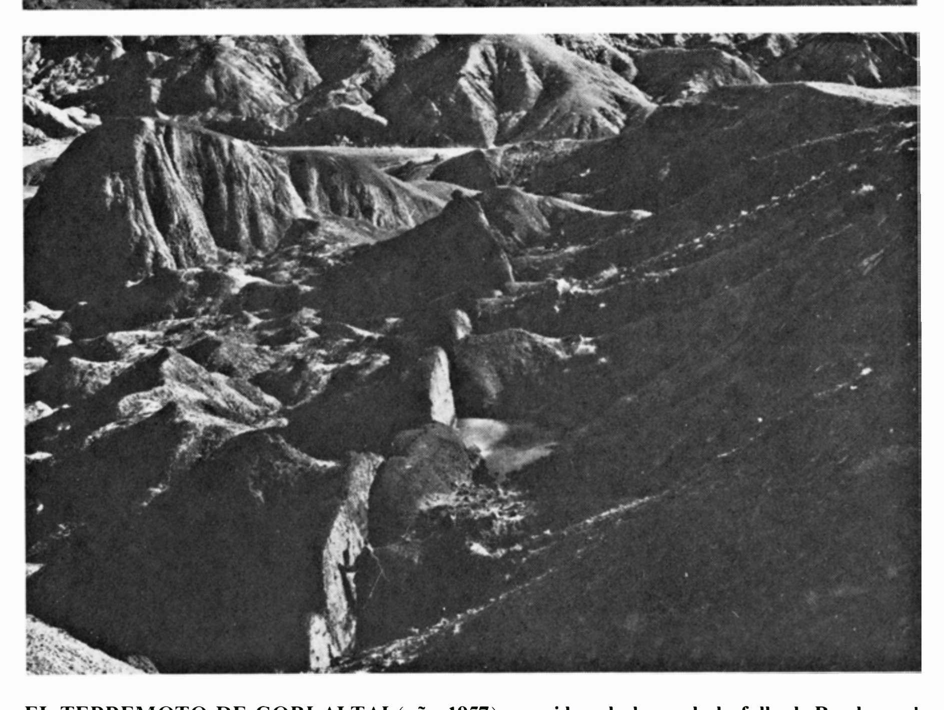
Los estudios de los movimientos sísmicos del Himalaya confirman las observaciones geológicas y demuestran que la India continúa hundiéndose por debajo de esta cordillera, según una dirección que varía entre norte y nordeste. Durante los últimos 50 años distintos investigadores han sugerido que el borde norte de la India sufría igualmente un hundimiento por debajo el altiplano del Tíbet y que, debido a ello, éste se encuentra a mayor altura.

Tanto nosotros como otros muchos geólogos creemos hoy en día que esto es muy poco probable, ya que no observamos ninguna prueba de que este hundimiento continúe en la actualidad y, mecánicamente, un fenómeno de este tipo parece ser del todo ideal.



UN ACORTAMIENTO Y UN ENGROSAMIENTO de la corteza parece haber ocurrido en la región de la cordillera de Tien Shan, al sudeste del lago Balkhash, según lo indica esta fotografía del ERTS centrada sobre los 40,5° de latitud norte y los 78,5° de longitud este. La foto muestra los sedimentos de la parte sur de Tien Shan plegados debido a la presencia de cabalgamientos que buzan hacia el norte, y están localizados bajo las formaciones sedimentarias de esta región. Los estudios sísmicos demuestran que la corteza es de 20 a 30 kilómetros más gruesa en el norte de Tien Shan que en zonas adyacentes que gozan de mayor estabilidad. Esto sugiere que en esta parte oriental de Tien Shan la corteza ha sufrido un acortamiento, o compresión, de hasta 300 kilómetros, quizás a causa del movimiento hacia el norte de la India.

Hace muy poco John F. Dewey y Kevin Burke, de la Universidad estatal de Nueva York, en Albany, han sugerido que el altiplano del Tíbet resulta de un acortamiento de la corteza, que se produjo en toda la región formada por el Himalaya y el Tíbet con el consiguiente comportamiento en acordeón de la corteza de este último, al contraerse horizontalmente y extenderse verticalmente. Si bien pensamos que este concepto puede aplicarse al Tíbet, no podemos aceptarlo por falta de pruebas convincentes. Por un lado, los estudios de los terremotos demuestran que la región no sufre acortamiento alguno sino una extensión en dirección este-oeste y, además, comparando las fotografías de satélite de esta última con las de otras zonas montañosas de Asia en donde se está produciendo acortamiento hoy en día, la deformación superficial del altiplano tibetano resulta ser mucho menor y, en el supuesto de que se haya producido plegamiento de la corteza, este fenómeno es menos reciente. Sin embargo, aun cuando consideremos que el espesor de la corteza del Tíbet ha llegado a duplicarse como resultado de la presión aplicada desde el sur, sólo pueden justificarse de este modo 600 o 700 de los 2000 kilómetros.



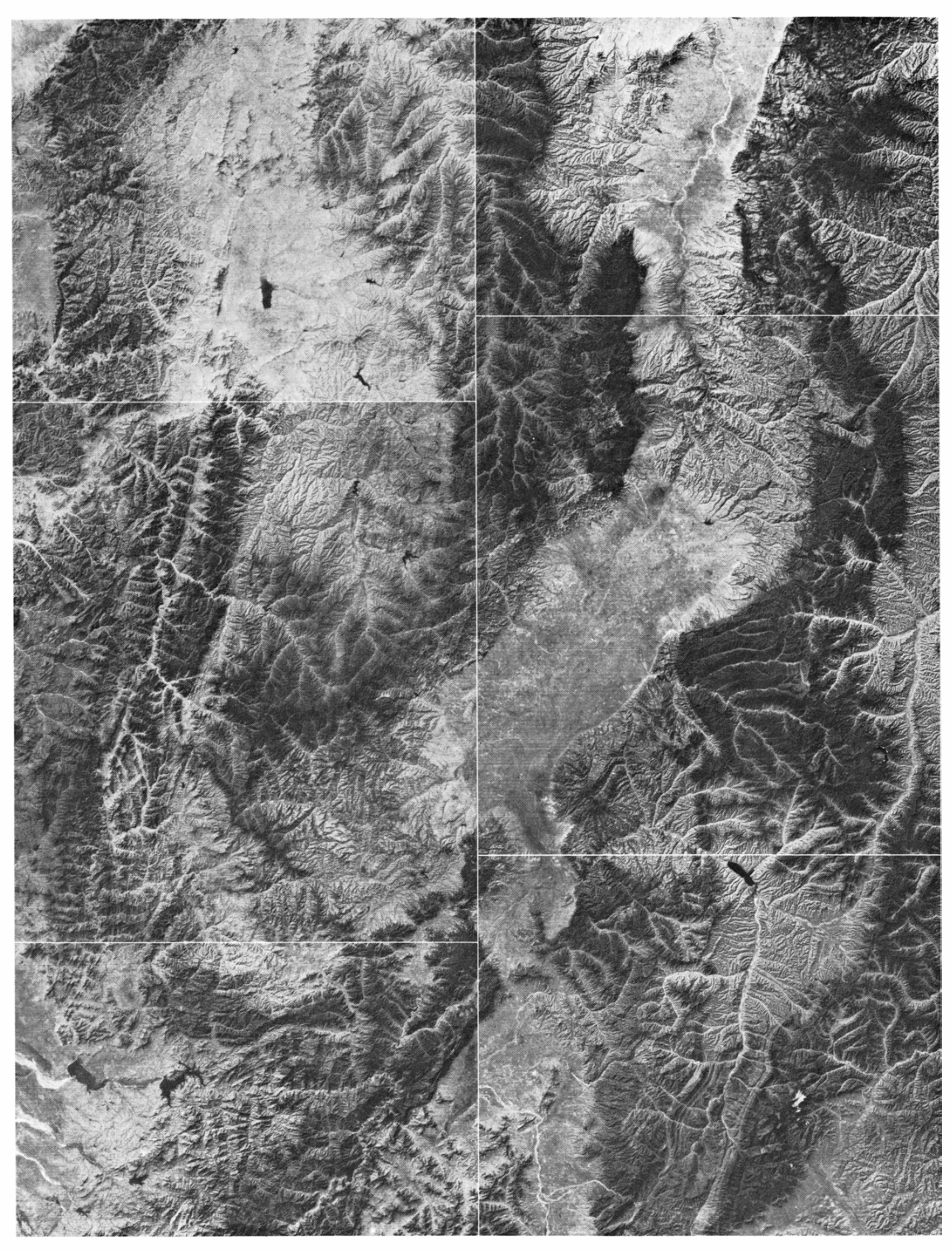
EL TERREMOTO DE GOBI-ALTAI (año 1957) ocurrido a lo largo de la falla de Bogdo en el desierto de Gobi, ocasionó los desplazamientos corticales observados en estas fotografías tomadas por el geólogo V. P. Solonenko. El terremoto, de magnitud 7,9 en la escala de Richter, produjo un movimiento lateral izquierdo, lo cual significa que para un observador situado sobre uno cualquiera de los dos labios de falla el desplazamiento es hacia su izquierda. Este movimiento alcanzó 10 metros y puede considerarse uno de los mayores desplazamientos habidos en una falla de rumbo

En la zona de Tien Shan, la distribución de la deformación está dominada también por fallas cabalgantes, con acortamiento en dirección norte. El sismólogo ruso V. I, Ulomov calcula que el valor del acortamiento en la parte occidental de la región es de unos 300 kilómetros, cifra a la que se llega imaginando lo que ocurriría si la corteza engrosada fuese aplanada (según sus palabras "con un rodillo de cocina"), con la consiguiente extensión lateral, hasta alcanzar su grosor normal. La mayoría de las demás soluciones en el plano de falla de los terremotos asiáticos indican un predominio general de las fallas de rumbo. Los cabalgamientos predominan únicamente en zonas retringidas, mientras que las fallas normales son poco frecuentes. Así, por ejemplo, en Mongolia, la mayoría de los terrenos están asociados a fallas de rumbo. De todos modos, los movimientos sísmicos de esta región presentan una orientación muy coherente del esfuerzo compresivo máximo en la dirección nordeste-sudoeste.

Aunque la disposición al este del Tíbet es más compleja, se parece a la disposición existente en Mongolia.

Así pues, la fracturación asociada a los terremotos indica que una gran parte de Asia sufre una compresión según una dirección intermedia norte-sur y nordeste-sudoeste, la cual es compatible con el movimiento de la India hacia el norte. La compresión hace que en determinadas partes de Asia la corteza sufra un acortamiento y aumente de espesor. Sin embargo, aquí nos encontramos de nuevo con que, incluso añadiendo unos hipotéticos 700 kilómetros de acortamiento cortical en el Tíbet a los 300 kilómetros correspondientes a la zona de Tien Shan, se pueden justificar solamente 1000 kilómetros de acortamiento, o sea, la mitad del movimiento total de la India hacia Eurasia, desde el comienzo de la colisión. Debe buscarse otra explicación para los restantes 1000 kilómetros de desplazamiento aun en el caso de que nuestras dudas acerca del acortamiento en "acordeón" ocurrido en el Tíbet sean erróneas.

Las fotos de satélite de Asia son elementos de primer orden que nos conducen a otra explicación respecto a los 2000 km de convergencia entre la India



PARTE DEL SISTEMA DE FOSAS DE SHANSI separa el extremo oriental del inaccesible altiplano de Ordos (comarca del desierto de Gobi) del fértil y densamente poblado valle del Huang Ho. Una fosa es una región hundida en donde los bloques corticales han sufrido una separación. Este mosaico formado con seis fotos del ERTS nos muestra los rasgos del relieve en una zona de 260 por 345 kilómetros cercana a Taiyuan, a unos 500 kilómetros al sudoeste de Pekín. Las fotografías de este mosaico fueron tomadas en diferentes épocas del año de modo que en algunas se observa nieve y hielo en el terreno y en los ríos. La fosa de Shansi y el sistema de fosas del Baikal se parecen a las grietas de tensión.

y Eurasia. Quizá los caracteres más notables observados en fotos aéreas de cualquier tipo sean las fallas de rumbo. Generalmente se presentan como rasgos lineales muy netos en la topografía. En Asia, este tipo de fracturas aparecen con una claridad pocas veces igualada en otras partes del mundo. Si bien algunas de ellas eran ya conocidas a partir de desplazamientos ocurridos durante los terremotos, o por estudios geológicos, algunas eran desconocidas hasta hace poco, por lo menos para la mayoría de geólogos.

La falla de Talasso Fergana, en la zona de Tien Shan, es bien conocida por los geólogos rusos y también se destaca claramente en las fotos sacadas por vía satélite. Los bruscos cambios de dirección de las crestas cercanas a la traza de la falla muestran nítidamente que el sentido del movimiento de ella es lateral derecho. La falla de Bolnai (o de Khangai), del norte de Mongolia, apareció en 1905 durante uno de los terremotos de mayor intensidad acaecidos en este siglo; a pesar de que es poco conocida fuera de Mongolia y la URSS, se distingue claramente en las fotos de satélite. Los detallados trabajos de campo realizados por los geólogos rusos S. D. Khilko, N. A. Florensov y sus colaboradores demuestran que un desplazamiento lateral izquierdo de unos pocos metros se extendió a lo largo de 370 kilómetros durante el terremoto de 1905. Quizá la falla de rumbo más larga y mayor del mundo sea la de Altyn Tagh, situada inmediatamente al norte del Tíbet. Según nuestras averiguaciones era desconocida como falla activa, al menos fuera de China, hasta que se pudo disponer de las fotos de satélite.

Para nosotros, la identificación de importantes fallas de rumbo en Asia fue un acontecimiento emocionante, ya que vimos en seguida y claramente la sencillez de su disposición. Combinando las soluciones en el plano de falla de los terremotos de Asia con los desplazamientos observados en superficie, después de haberse producido alguno de los movimientos sísmicos más notables, y con los caracteres tectónicos asociados a las fracturas visibles en las fotos de satélite, hemos podido determinar el sentido del movimiento inherente a las fallas principales. En China y Mongolia, las fallas de rumbo tienen una dirección general aproximada este-oeste y responden a un movimiento lateral izquierdo. Las fallas de la región de Tien Shan y algunas de Mongolia, o de sus cercanías, tienen una dirección norte-sur o noroeste-sudeste y responden a un movimiento lateral derecho. No es otra la configuración que cabría esperar si la región se hubiera comprimido según una dirección comprendida entre la norte-sur y la nordeste-sudoeste, siendo seguramente esta compresión la consecuencia probable de la colisión entre la India y Eurasia. Además, el predominio del movimiento lateral sugiere que el movimiento de la India hacia el norte se ve facilitado esencialmente gracias a la desaparición del material situado en medio de los dos continentes en colisión.

En general, las fallas más claras, y, por tanto, presuntamente las más importantes, son las de desplazamiento levógiro, que siguen una dirección este-oeste. Así pues, la mayor parte del desplazamiento lateral de China, al dejar paso al avance de la India, sería hacia el este. Ello resulta fácilmente comprensible: el material desplazado hacia el oeste tendría que empujar contra millares de kilómetros de masa continental eurasiática.

El movimiento de China hacia el este encuentra con facilidad espacio suficiente cabalgando sobre las placas oceánicas marginales del Pacífico. Un ejemplo de la vida cotidiana para explicar este tipo de desplazamiento lo tenemos cuando exprimimos un tubo de pasta dentífrica. Los dedos de la mano corresponden a la India y resto de Eurasia, la terminación cerrada del tubo equivale a la vasta región continental de Eurasia que, básicamente, bloquea el movimiento a gran escala hacia el oeste y, por último, la boca del tubo semeja las zonas de subducción del Pacífico occidental, haciendo China y Mongolia las veces de pasta dentífrica.

La existencia de importantes fallas de rumbo nos ayuda a explicar igualmente algunas otras peculiaridades evidentes de la geología de Asia. Si bien la mayor parte de este continente experimenta una compresión horizontal según una dirección intermedia norte-nordeste, existen dos excepciones notables. El lago Baikal ocupa una parte del sistema de fosas tectónicas del mismo nombre que representan una extensión noroeste-sudeste. De forma parecida, Deng Qidong y sus colaboradores, del Instituto Geológico de Pekín, describen el *graben* de Shansi en la China oriental como un sistema de fosas originado por otra extensión de igual orientación. Interpretamos ambos sistemas como resultado parcial de su proximidad a las fallas de rumbo, pudiéndose comparar con las grietas de tensión que se desarrollan en los extremos de las grietas de cizalla.

Según esta interpretación, después de la colisión de la India contra Eurasia disminuyó la velocidad relativa del movimiento, pero la primera continuó su movimiento hacia la segunda a una velocidad de cinco centímetros por año. Al hacerlo provocó una compresión hacia arriba de sus relieves frontales y, lo que es más importante, desplazó los materiales que le impedían el paso. El flujo de estos materiales pudo haber abierto fosas tectónicas según una dirección perpendicular a la de acortamiento de forma que, en cierto modo, la India se ha metido como una cuña que ha seccionado en dos a Eurasia. La distribución global de la deformación es físicamente parecida a la provocada por la indentación en un medio plástico mediante un instrumento rígido. Creemos que las soluciones propuestas por los ingenieros mecánicos para resolver problemas de indentación pueden permitirnos tratar cuantitativamente el estado de esfuerzo a que está sometida Asia, y puede conducirnos a una mejor comprensión de la deformación a gran escala de los continentes.

La suposición básica y no demostrada en el esquema descrito es que se han producido grandes desplazamientos horizontales en las fallas de rumbo de Asia. Hasta donde sabemos, la falla estudiada de forma más cuidadosa entre todas las mencionadas es la de Talasso Fergana, en la URSS. No hay acuerdo entre los geólogos rusos acerca de la magnitud del desplazamiento que ha tenido lugar sobre ella, pero V. S. Burtman sugiere que puede ser de 200 a 250 kilómetros de movimiento lateral derecho. Indica, sin embargo, que la mayor parte del desplazamiento se produjo antes de la colisión entre la India y Eurasia, no pudiéndose estimar la magnitud del mismo después de este acontecimiento. En relación a las otras fallas podemos decir solamente que la mayoría destaca en las fotos de satélite tanto como la falla de San Andrés, en California, la cual ha sufrido un desplazamiento lateral derecho de 300 kilómetros en los últimos 25 millones de años. Si los desplazamientos, tan nitidos a lo largo de estas fallas asiáticas, son comparables a los producidos en la de San Andrés, tal como sospechamos, se puede concluir que la mayor parte de los 2000 kilómetros de convergencia entre la India y Eurasia pueden atribuirse al desplazamiento lateral de China. Por otra parte nuestra hipótesis se desacreditaría fatalmente si se pudiera demostrar que el desplazamiento a lo largo de las fallas

principales ha sido de sólo unos pocos kilómetros o, como máximo, de unas decenas, a lo largo de los últimos 40 millones de años.

De todos modos es virtualmente cierto que las fallas de rumbo desempeñan un papel clave en el proceso de suturación entre continentes. Hace algunos años McKenzie apuntó que en Oriente Medio, la otra región de importancia en cuanto a colisión activa de continentes se refiere, el movimiento del subcontinente arábigo hacia Eurasia obliga a parte de Turquía a moverse hacia el oeste en dirección perpendicular a la de los bloques continentales en convergencia. Queda claro, además, gracias a los detallados estudios de Mattauer y sus colaboradores en determinadas partes de algunas cordilleras de montañas antiguas, como las existentes en Francia, España y Marruecos, que el movimiento a lo largo de las fallas de rumbo fue importante hasta mucho después de la colisión continental. Es muy probable que, una vez estudiadas otras cadenas de plegamiento antiguas, se encuentren pruebas de movimientos horizontales a gran escala a lo largo de las fallas de rumbo. Este tipo de análisis sobre tectónica continental dista mucho de ser una aplicación directa de la tectónica de placas a los continentes. La aplicación de la tectónica de placas a la deformación de Asia requeriría tantas placas que el concepto perdería toda su fuerza. En el futuro, esperamos, esto será igualmente aplicable a las colisiones continentales más antiguas.

Entendemos la tectónica de Asia como una consecuencia directa de los movimientos de las placas. Los terremotos y las grandes fallas de la India, China, Mongolia y la URSS pueden atribuirse a un simple fenómeno: el movimiento hacia el norte del subcontinente índico que hace avanzar la placa de la India hacia la placa de Eurasia. Quizá lo más interesante de esta interpretación resida en que indica que el movimiento de la India ocasionó la deformación de una región situada a más de 3000 kilómetros de distancia. Puesto que las montañas se formaron en virtud del movimiento habido a lo largo de las fallas, y dado que el clima de estas regiones está a su vez profundamente influido por la topografía, las condiciones ambientales de una gran parte de Asia, como, por ejemplo, los climas extremados del Himalaya, altiplano del Tíbet, desierto de Gobi, Mongolia y partes de China, pueden atribuirse a una colisión que ha continuado durante 40 millones de años.

El interferón

La proteína que defiende a las células del ataque de los virus se descubrió hace ya veinte años. Tras el esfuerzo por lograr su purificación y seguir su modo de actuación, se investiga ahora su capacidad antivírica general

Derek C. Burke

Han pasado 20 años desde que Alick Isaacs y Jean Lindenmann descubrieron una sustancia que liberan las células expuestas a un virus y que, a renglón seguido, protege a otras células de las infecciones víricas. A esta sustancia le dieron el nombre de interferón, porque evidentemente parecía ser el agente de la interferencia vírica, fenómeno que estaban estudiando en el National Institute for Medical Research de Londres. Desde los años treinta, se había observado que la infección por virus, de un cultivo de células animales o de un animal de laboratorio, "interfería" durante algún tiempo la infección ulterior por otros virus. Isaacs y Lindenmann hallaron que la acción del primer virus no se limitaba al simple bloqueo del acceso de los otros virus a las mismas células, sino que, además, inducía en las células infectadas la liberación de alguna sustancia que se difundía en el medio; y así, si se extraían las células y se incorporaba el medio a otro cultivo, las nuevas células se volvían resistentes a los virus.

Isaacs y varios de los que entonces trabajábamos en su laboratorio no tardamos mucho en deducir algunas propiedades del interferón. Al parecer, se trataba de una proteína, de peso molecular relativamente bajo. Su acción no era específica respecto de los virus. Es decir, que no reaccionaba selectivamente con un virus determinado, tal como reacciona un anticuerpo con su antígeno, ni tampoco reaccionaba con una partícula vírica libre, tal como reacciona un anticuerpo con una sustancia antigénica. En cambio, reaccionaba con las células, y las células afectadas se volvían resistentes a un gran número de virus.

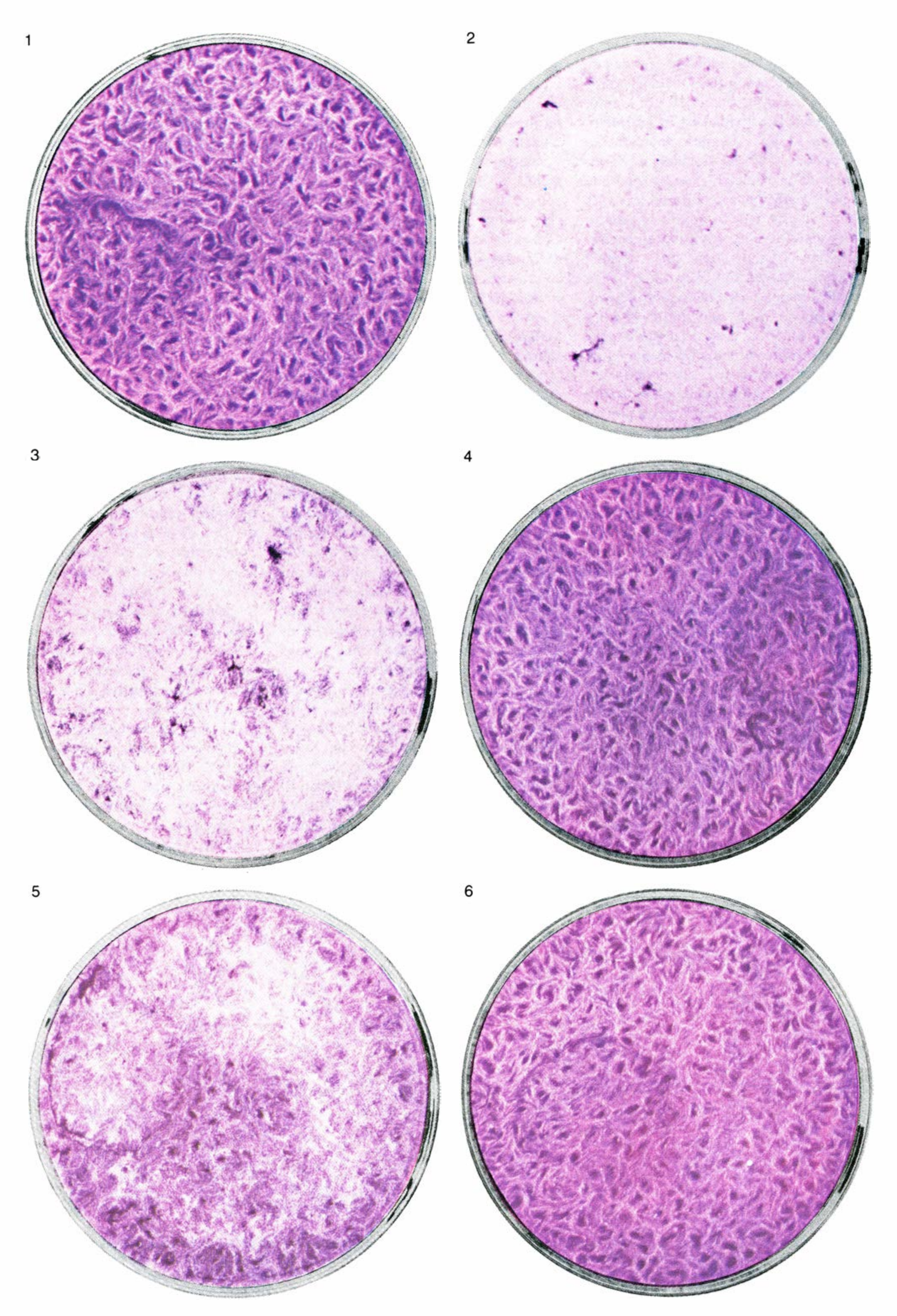
Por otra parte, el interferón solía ser específico respecto de la especie: el interferón fabricado por células de polluelo era activo en otras células del mismo animal, pero no en células de pato, etc. En otras palabras, el interferón era un agente celular que, al parecer, integraba la primera línea de defensa natural frente a los virus en las células animales. Debido a que era liberado por las células (y, por tanto, podía recogerse con facilidad), a que era eficaz para combatir muchos virus y a que era un producto celular natural con poca probabilidad de lesionar a las células o de provocar resistencia en los virus sometidos a su acción, el interferón prometía tener un gran valor clínico. Se suponía que, si se conseguía aislarlo, podría administrarse para prevenir o combatir la infección vírica, de una forma muy similar a la utilización de los antibióticos en la lucha contra la infección bacteriana.

Han pasado dos décadas y aún no se ha cumplido esa promesa. Se ha invertido mucho tiempo en conocer qué es exactamente el interferón, cómo se fabrica y su funcionamiento en el interior de la célula. Por todo ello, y porque resulta difícil preparar grandes cantidades del mismo, cuesta establecer la eficacia exacta que podría tener dicha sustancia en medicina clínica. No obstante, a lo largo de los últimos años, se han preparado pequeñas cantidades de lo que parece ser interferón puro, de modo que ya se puede proceder a la tarea de describirlo y analizarlo. Algo sabemos en la actualidad de la forma en que es inducido y de su manera de interferir la infección vírica. Por otra parte, son varias las pruebas clínicas que han puesto de manifiesto que el interferón podría intervenir contra ciertas enfermedades víricas graves.

Dado que el interferón es una proteína, en teoría, tendría que haber sido sólo cuestión de tiempo conseguir su purificación a través de los procedimientos típicos para tales moléculas: "salado", cromatografía de intercambio iónico, electroforesis sobre gel, etc. Varios investigadores intentaron la purificación por estos métodos. El resultado fue desalentador: aunque los métodos típicos lograban la purificación de la sustancia hasta un grado considerable, el producto final no era una entidad homogénea que pudiera describirse como una proteína única. El problema principal en el estudio del interferón ha radicado en la dificultad para aislar la proteína pura.

Una de las razones que se aducen para explicar tal dificultad es la actividad específica, muy elevada, del interferón.

EL EFECTO DEL INTERFERON se mide determinando su capacidad para proteger las células. Las fotografías indican el grado de purificación del interferón de fibroblastos humanos obtenido por Ernest Knight, Jr., de E. I. du Pont de Nemours & Company. En una cápsula de Petri se hace crecer una sola capa de fibroblastos humanos *(1)*. Se añade al medio el virus de la estomatitis vesicular; a las 24 horas, las células han muerto *(2)*. Si al cultivo se le aplica, durante 24 horas antes de ser sometido al desafío vírico, interferón no purificado a una concentración de unos 0,02 microgramos por mililitro, la mitad de las células aproximadamente quedan protegidas *(3)*; en otras palabras, esa concentración de interferón crudo nos da una unidad de actividad biológica. Todas las células quedan protegidas con diez unidades biológicas del mismo material *(4)*. Si se aplica el interferón purificado, se obtiene el mismo grado de protección con cantidades mucho menores: 0,005 nanogramos por mililitro para dar una unidad de actividad, o una protección del 50 por ciento *(5)*, y 0,05 nanogramos para proteger a todas las células *(6)*. Se ha purificado unas 4000 veces; la actividad de la proteína es de unos 200 millones de unidades por miligramo.


1
2
3
4
5
6

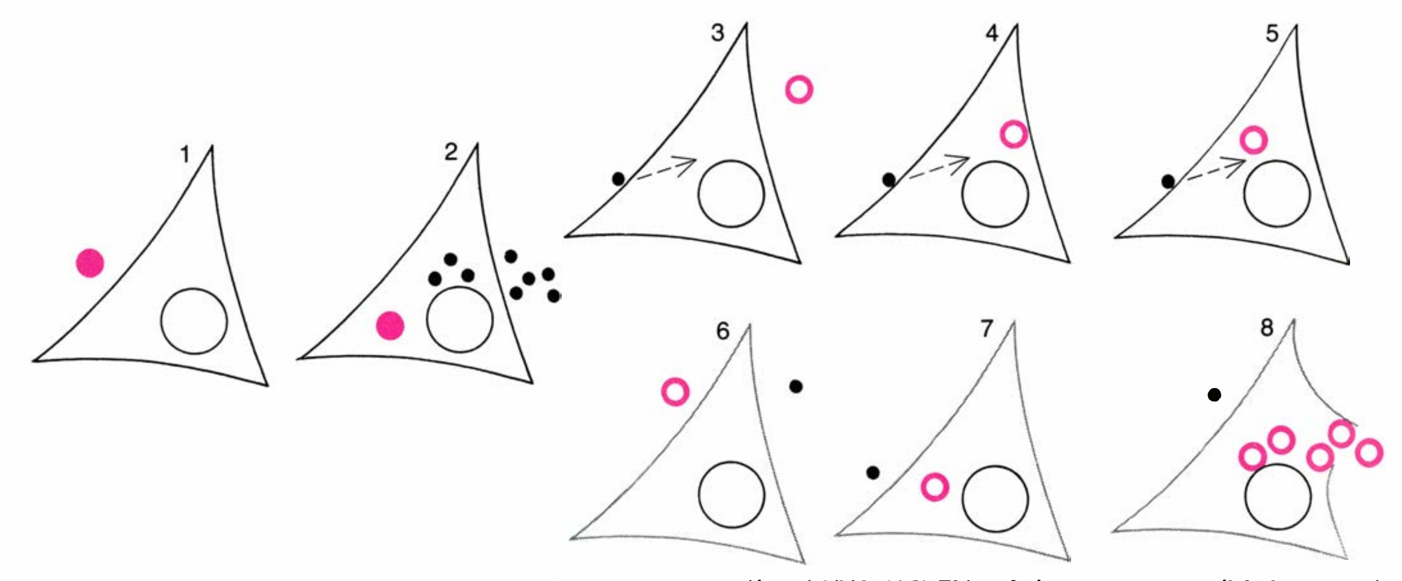
La actividad se mide en unidades arbitrarias, siendo una unidad la cantidad de interferón necesaria para reducir la producción del virus en un 50 por ciento. Al parecer, la actividad específica del interferón puro es de aproximadamente mil millones (10^9) de unidades por miligramo de proteína. El material del que se parte para realizar la purificación, es decir, el extracto crudo de interferón procedente de un cultivo de células tratadas con un virus, tiene una actividad característica que sólo alcanza de 1000 a 10.000 unidades por miligramo de proteína, por lo que es muy grande el volumen de material crudo necesario para obtener una cantidad muy pequeña de interferón purificado. La preparación de volúmenes tan grandes resulta costosa y cada paso de la purificación precisa industriosos procedimientos de ensayo.

Existe una dificultad fundamental: las moléculas de interferón tienden a adherirse a otras moléculas –en particular a otras moléculas proteicas– presentes en el extracto crudo. Como consecuencia, alguna vez se ha dado el caso de que muchos investigadores, incluyéndome a mí mismo, han purificado una proteína que era distinta de la deseada, a la que se había unido una ligera cantidad de interferón.

En los últimos tiempos, dichas dificultades han sido ampliamente superadas por el desarrollo de dos técnicas nuevas. La primera es la cromatografía de afinidad *(véase la ilustración superior de la página 33)*. Se empaqueta una columna de cromatografía con una matriz sólida a la que está unida una sustancia respecto de la cual el interferón tiene gran afinidad. Cuando el extracto crudo de interferón atraviesa la columna, tras haber sido vertido en su parte superior, el interferón queda retenido sobre las partículas de la matriz, mientras que muchas impurezas atraviesan la columna y salen por el otro extremo. Posteriormente, el interferón se eluye o se lava con un disolvente adecuado. Kurt Paucker y sus colaboradores, del Medical College de Pennsylvania, han conseguido un alto grado de purificación utilizando como sustancia de alta afinidad el anticuerpo del interferón. William A. Carter y sus colaboradores, del Roswell Park Memorial Institute de Buffalo, aprovecharon una de las interacciones con otras proteínas, que suelen ser fuente de problemas (en este caso, con la albúmina) para lograr una purificación de 2000 veces en un solo paso. Otro tipo de cromatografía de afinidad explota el hecho de que el interferón es una glucoproteína, es decir, una molécula que tiene una o más cadenas laterales hidrocarbonadas, que terminan en un grupo de ácido siálico. Dicho grupo se une de forma muy específica a las sustancias vegetales denominadas lectinas. La mencionada interacción fue aprovechada primero por Rudolf Wein y sus colaboradores, de los laboratorios Sandoz de Viena, y, más tarde, por otros investigadores.

La segunda técnica que ha contribuido a la purificación del interferón se desarrolló a partir del descubrimiento de cómo bloquear la interacción del interferón con otras proteínas, sin destruir la actividad biológica de aquél. William E. Stewart, de la Universidad Católica de Lovaina (Bélgica), hirvió interferón preparado a partir de leucocitos humanos (glóbulos blancos de la sangre), en dodecilsulfato sódico, que es un fuerte detergente. Dicho tratamiento desnaturaliza las proteínas, es decir, despliega la larga secuencia de aminoácidos de la molécula proteica y la vuelve inactiva. Stewart comprobó que si se dejaba enfriar y se separaba el detergente, el interferón recuperaba su actividad biológica original: se renaturalizaba. Esta estabilidad del interferón, en fuertes condiciones de desnaturalización que destruían las interacciones ordinarias entre proteínas, permitió separar las moléculas de interferón del interferón retenido por adsorción sobre otras especies moleculares.

Stewart también comprobó que, si el preparado de interferón se hervía en el detergente en presencia de un agente reductor, sólo se podía renaturalizar parcialmente *(véase la ilustración inferior de la página opuesta)*. Sugirió que esto se debía a que el extracto crudo contenía dos tipos de interferón que se distinguían porque uno de ellos no era estable en condiciones reductoras. De esta forma se posibilitó la separación de dos tipos de interferón de leucocitos humanos. El componente principal, denominado *Le*, tiene un peso molecular de 15.000; no se renaturaliza tras un tratamiento con detergente en condiciones reductoras. El otro componente, que constituye una proporción menor del interferón total, se denomina *F*; tiene un peso molecular de aproximadamente 21.000 y es activo tanto si hay un agente reductor presente en el tratamiento con detergente como si no lo hay. Ambos interferones difieren por su grado de actividad antivírica en células de otras especies; el efecto protec-



EL INTERFERON *(puntos negros)* **producido por una célula infectada por un virus** *(disco de color)* **se libera y difunde en el medio** *(1, 2)*. **De alguna manera, el interferón protege a otra célula de la misma especie** *(3)* **ante un virus diferente** *(círculo abierto de color)* **cuya proliferación es inhibida** *(4,5)*. **El interferón no protege a una célula de una especie distinta** *(6)* **en el interior de la cual se puede multiplicar el virus y termina matando a la célula** *(7, 8)*. **Esta representación de los efectos del interferón refleja el estado de conocimiento que se tenía hacia 1960.**

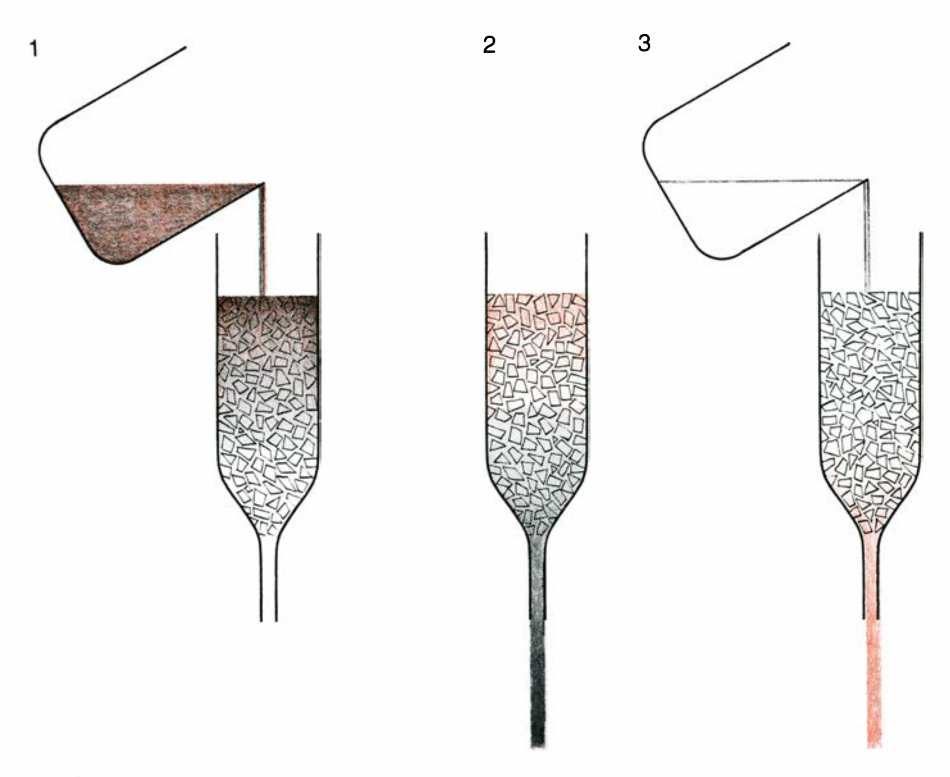
tor del tipo *Le* en células de conejo es tan sólo alrededor de un 5 por ciento de su efecto en células humanas, mientras que el tipo *F* es igualmente eficaz en ambos tipos de células. El tipo *F* es una glucoproteína; cabe que el tipo *Le* no lo sea. También se ha demostrado que el tipo *F* es muy similar al interferón producido por fibroblastos humanos, es decir, por los precursores de las células del tejido conjuntivo.

El tipo *Le* puede separarse, a su vez, en dos proteínas cuyas moléculas tienen un tamaño bastante similar. Nos encontramos, así pues, con una situación interesante: los fibroblastos humanos elaboran un solo tipo de interferón, mientras que los leucocitos humanos fabrican dos tipos y trazas de un tercero. ¡No es de extrañar que la purificación haya presentado tantas dificultades! No sabemos cuál es la diferencia química entre estos tipos de interferón, pero es probable que sus moléculas tengan secuencias de aminoácidos distintas, lo que sugiere una de dos causas: que estén codificados por genes diferentes o que el producto de un mismo gen sea procesado de distinta manera en ambos tipos de células.

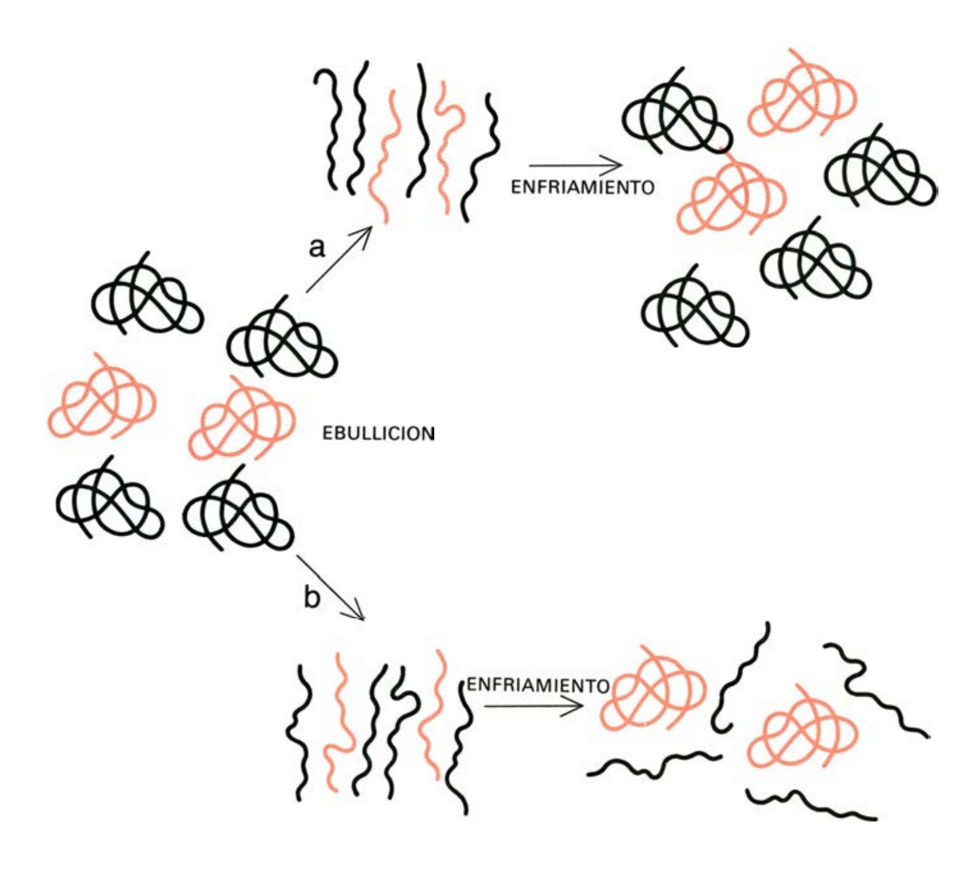
Explotando al máximo las técnicas de cromatografía de afinidad y de tratamiento con detergente, varios investigadores han conseguido notables progresos en la purificación. Por ejemplo, Ernest Knight, Jr., del departamento central de investigación y desarrollo del E. I. du Pont de Nemours and Company, ha obtenido interferón de fibroblastos humanos en una forma que, según él, sería homogénea. Se trata de una proteína más bien pequeña, cuyo peso molecular es 20.000. Si se consigue acumular una cantidad suficiente de este interferón, se podrá proceder a la determinación de su secuencia de aminoácidos; a partir de la cual, la síntesis artificial de la proteína sería algo muy posible de realizar.

Pero mientras no se lleve a cabo dicha síntesis, dependemos de las células para la fabricación de interferón. Existen dos formas de solucionar el problema de la elaboración. Se puede inducir la producción del interferón en las células de un cultivo de tejido y luego recoger la proteína, purificarla y administrarla a los pacientes. O bien, se puede administrar directamente a los pacientes un inductor que estimule la fabricación de interferón endógeno por las propias células del paciente. En cualquier caso, hay que saber cuál es la mejor manera de inducir la producción de interferón en las células.

El interferón se descubrió por primera vez al tratar células de polluelo con el



CROMATOGRAFIA DE AFINIDAD, uno de los métodos para purificar el interferón. El interferón crudo se vierte en una columna que contiene una matriz sólida (un polvo grueso) a la que está ligada una molécula que se une fuertemente al interferón *(1)*. Al unirse a la molécula, el interferón queda retenido en la columna, mientras que las impurezas del extracto crudo tienden a atravesarla *(2)*. Se eluye el interferón *(en color)* haciendo pasar un disolvente a través de la columna *(3)*.

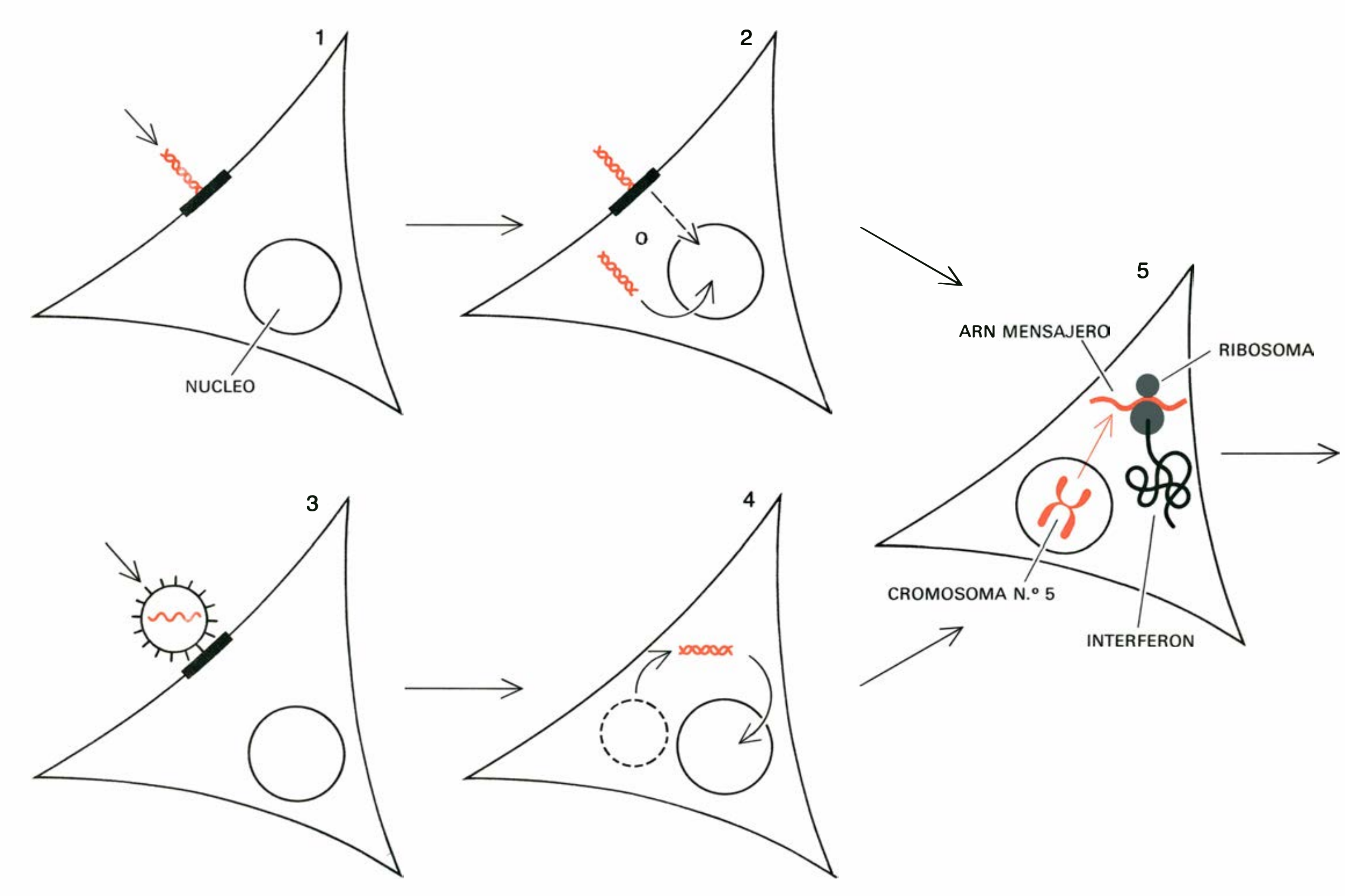


LA EBULLICION CON EL DETERGENTE dodecilsulfato sódico *(a)* desnaturaliza al interferón de leucocitos (glóbulos blancos) humanos, es decir, despliega su cadena proteica e inactiva a la molécula; si se deja enfriar, la proteína se renaturaliza y recupera su actividad. Si se enfría después de hervirla con el detergente, y en presencia de un agente reductor, 2-mercaptoetanol *(b)*, resulta que sólo se renaturaliza parcialmente el interferón. Al parecer, el interferón de leucocitos tiene dos componentes, uno de los cuales *(negro)* no se renaturaliza después de la ebullición en condiciones reductoras. El tratamiento por ebullición destruye la mayoría de las interacciones entre proteínas; separa, pues, el interferón de algunos contaminantes, seguir puede apreciarse.

virus de la gripe, pero muy pronto quedó de manifiesto que otros muchos virus eran capaces de inducir la producción de la sustancia. Ni siquiera se exigía que el virus fuera virulento; bastaba con virus atenuados. Luego, se halló que también eran inductores de interferón otros agentes distintos de los virus. Los primeros agentes de este tipo que se citaron fueron dos antibióticos, la helenina y el estatolón, que habían sido extraídos de mohos. Los investigadores de los laboratorios de investigación de la Merck Sharp & Dohme emprendieron la búsqueda del principio activo de la helenina. Pero, ya anteriormente, habían presentado un trabajo indicando que las respuestas inmunes de los animales eran reforzadas por los polinucleótidos sintéticos: análogos del ADN o del ARN que se sintetizan en el laboratorio por unión arbitraria de las subunidades o nucleótidos de los ácidos nucleicos. Observaron que el poli *rI:rC*, un ARN sintético que consiste en cadenas apareadas formadas por subunidades de ácido riboinosínico y ácido ribocitidílico, respectivamente, era muy activo como inductor del interferón. Posteriormente hallaron que el principio activo de la helenina también era un ARN de doble cadena, y demostraron que procedía de un virus; los mismos hechos se confirmaron en el caso del estatolón.

Los investigadores de la Merck Sharp & Dohme continuaron estudiando varios ácidos nucleicos sintéticos y naturales, inyectándolos en conejos y comprobando la capacidad del suero sanguíneo del conejo para proteger a un cultivo celular ante una infección vírica. Se descubrió que varios ARN de cadena doble eran eficaces inductores, siendo particularmente activo el poli *rI:rC*. El ADN no era activo, ni tampoco lo era el ARN de cadena sencilla; los investigadores sugirieron que la cadena sencilla de ARN, que constituye el material genético de la mayoría de los virus animales, debe replicarse (es decir, formar una réplica de sí mismo) en el interior de la célula para producir un ARN de doble cadena, antes de que pueda desencadenar la producción de interferón.

Una investigación más profunda demostró que los virus y los ARN de doble cadena no son las únicas sustancias capaces de inducir la formación de interferón (aunque sean las únicas activas en las células de los cultivos de tejido característicos). En especial, se descubrió que las células linfoides, que fabrican anticuerpos y participan en la inmunidad celular, elaboran interferón al ser tratadas con el antígeno adecuado, mecanismo que quizá desempeñe una función en el sistema inmune. El interferón inducido por un antígeno es fácil de distinguir, tanto inmunológicamente como por sus propiedades físicas, del interferón producido por células humanas como respuesta a los virus o a los ARN. Esto confirma la impresión de que existen varios interferones que tienen en común una actividad antivírica, pero difieren entre sí por otros aspectos. Se ha inducido la producción de interferones en todo el animal mediante una variedad de sustancias. El análisis genético de la producción de interferón en ratones endogámicos demuestra que los interferones fabricados como respuesta a distintos inductores están controlados por distintos genes del huésped –otro indicio de la complejidad de tal sistema.



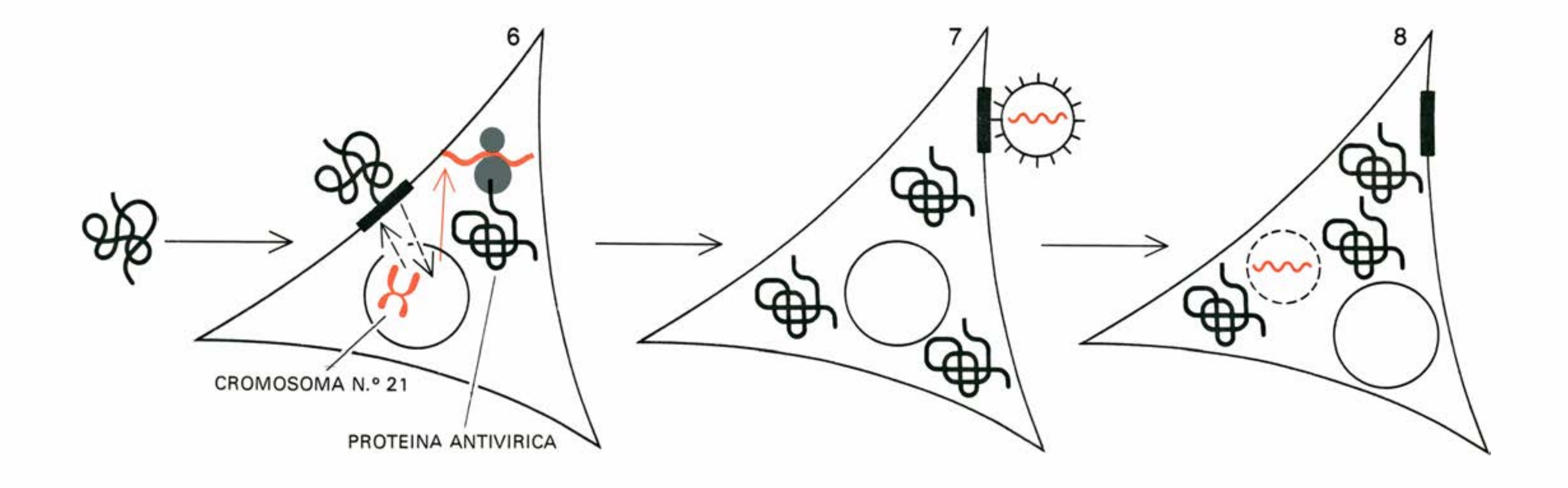
LA INDUCCION Y EL MECANISMO DE ACCION del interferón son procesos que se conocen con bastante detalle. Se puede inducir a una célula a elaborar interferón mediante un ARN de doble cadena, que se une a un receptor de la superficie celular *(1)* y luego penetra en la célula o bien envía un mensaje desde la superficie al núcleo celular *(2)*. También se puede inducir la síntesis de interferón mediante un virus, que se une a un receptor *(3)* y luego penetra en la célula, y en su interior fabrica ARN nuevo (que quizá tenga que ser de doble cadena).

La manera de realizar la inducción es una cuestión más fácil de estudiar en células de un cultivo de tejido que en un animal (dado que así se puede estudiar una sola población de células) y utilizando ARN de doble cadena como inductores. Parece estar claro que estos ARN interactúan con un receptor situado sobre la superficie de la célula. Ello explicaría por qué cualquier inductor nucleotídico debe reunir unas condiciones estructurales muy concretas: debe ser de doble cadena; ambas cadenas deben presentar el anillo de ribosa típico del ARN, que tiene un grupo hidroxilo (OH) en el átomo de carbono que está en posición 2; su temperatura de fusión debe ser superior a los 60 grados Celsius, es decir, que debe estar en la forma helicoidal –no desenrollada– a la temperatura del experimento de inducción, y su peso molecular debe estar por encima de un valor mínimo.

Cuando el experimento se lleva a cabo a cuatro grados Celsius, hay una interacción entre el ARN y el receptor; pero, al parecer, la unión es tan sólo superficial, ya que un tratamiento con el enzima ribonucleasa, que digiere al ARN, impide la formación del interferón. Si el preparado se calienta hasta 37 grados (la temperatura corporal), la ribonucleasa pierde su capacidad de influencia en la formación del interferón. Esto tiene dos causas posibles: o bien el ARN ha penetrado en la célula a esa temperatura, y, por tanto, ya no está sometido al ataque del enzima, o bien, a esta temperatura superior, el ARN puede enviar un mensajero al interior de la célula, sin penetrar en ella el propio ARN, dando así la señal que desencadena el proceso de inducción. No estamos todavía en condiciones de poder distinguir entre ambas posibilidades; se sabe que los ARN de doble cadena pueden penetrar en las células, pero ningún experimento ha demostrado aún que tengan que hacerlo. Por otra parte, está claro que, cuando un virus induce la producción de interferón, penetra realmente en la célula y sintetiza ARN vírico nuevo en el interior de ésta; hasta los virus inactivados que inducen la producción de interferón actúan de esta forma. Como ya indiqué anteriormente, quizás eso se deba a que el ARN vírico de cadena sencilla debe replicarse y formar un ARN de doble cadena que sirve entonces de inductor.

¿Qué ocurre luego, una vez el inductor ha penetrado en la célula o ha enviado un mensaje hacia su interior?

La especificidad del interferón respecto de la especie sugería que su formación es controlada por la célula huésped, lo cual quedó confirmado por el hecho de que la formación de interferón es inhibida por la actinomicina D (lo que significa que para dicha formación es necesaria la síntesis de ARN sobre un molde de ADN, presumiblemente perteneciente al ADN de la célula huésped) y también por la cicloheximida (lo que significa que también es necesaria la síntesis de proteínas). Parece, pues, que el inductor desencadena de alguna manera la "desrepresión" de un gen de la célula huésped, tras lo cual dicho gen es transcrito a ARN mensajero, que, a su vez, se traduce a una proteína. Dicho esquema se ha podido confirmar de diversas maneras. Y. H. Tan y Frank H. Ruddle, de la Universidad de Yale, han



que sirve como inductor *(4)*. En cualquier caso, la sustancia inductora provoca la síntesis, a partir del cromosoma humano número 5, de un ARN mensajero que se traduce a la proteína interferón *(5)*. El interferón sale de la célula y se une a otra célula humana *(6)* por medio de un receptor codificado por el cromosoma número 21. De este modo se induce a este mismo cromosoma o a otro a sintetizar ARNm, que se traduce a una proteína antivírica, protectora de la célula: si aparece otro virus y se une a la célula *(7)*, aunque penetre en ella *(8)*, no puede multiplicarse.

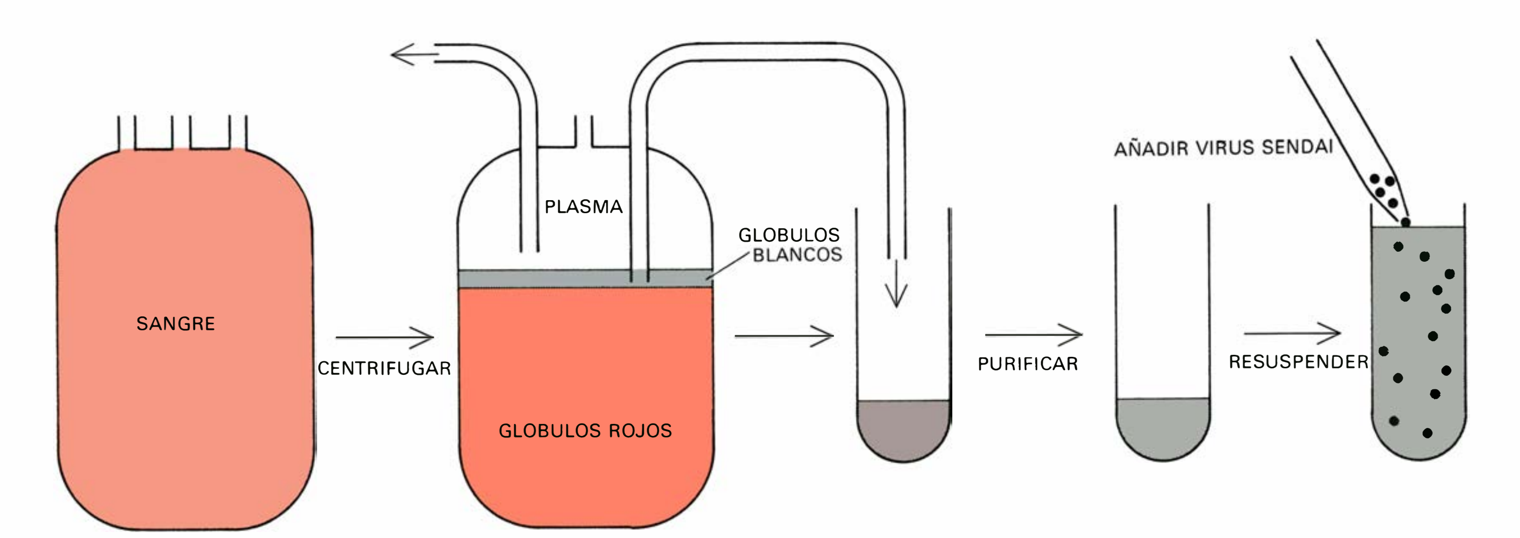
demostrado, por medio de unos experimentos con células híbridas de ratón y humanas, que el cromosoma humano número 5 es esencial para la formación de interferón. La hipótesis de un control de la síntesis de interferón por genes localizados en el núcleo quedó respaldada por un descubrimiento que yo mismo hice, a saber, que la eliminación del núcleo de la célula huésped impide la formación de interferón.

La vía desde el ARN hasta la proteína también quedó demostrada por la extracción de ARN mensajero a partir de células productoras de interferón y la traducción posterior de este ARN hasta la formación de interferón. Esta traducción fue llevada a cabo por primera vez en células pertenecientes a una especie diferente de la que había formado el ARN mensajero, realizando el experimento Jacqueline De Maeyer-Guinard, Edward De Maeyer y Luc Montagnier de la Fondation Curie-Institut du Radium, de Orsay, Francia. Añadieron ARN, procedente de células de ratón productoras de interferón, a células de polluelo a las que se inducía así a formar interferón de ratón. Más tarde, Paula M. Pitha y sus colaboradores de la Johns Hopkins School of Medicine comprobaron que cuando se inyectaba el mensajero de interferón de ratón en oocitos (células precursoras de los óvulos) de la rana *Xenopus*, también se realizaba la traducción. Por fin, el mensajero del interferón fue traducido en un sistema de síntesis proteica libre de células (una mezcla de extractos celulares en tubo de ensayo); el interferón producido *in vitro* quedó identificado no sólo por su actividad biológica sino también por precipitación con un anticuerpo específico del interferón, formándose un polipéptido del peso molecular esperado. Por otra parte, estos últimos resultados implican que el interferón no depende, para su actividad biológica, de sus cadenas laterales hidrocarbonadas, dado que en un sistema libre de células estas cadenas no pueden unirse a la molécula. Sin emba go, todos estos experimentos no contestan al interrogante principal: ¿cómo activa el inductor al gen de la célula huésped para que fabrique ARNm de interferón?

Una vez sintetizado el interferón, ¿cómo actúa en otras células para defenderlas de los virus? En primer lugar, debe interactuar con las otras células. Al parecer, esta interacción implica la unión a un receptor de la superficie celular, y también en este caso parece que la unión es inicialmente superficial: el interferón se une al receptor a cuatro grados, pero a esa temperatura tan baja su efecto antivírico puede ser destruido al tratar las células con enzimas que digieren las proteínas. El efecto antivírico se logra si se calienta el sistema hasta 37 grados, temperatura a la cual se vuelve insensible a la acción digestiva de los enzimas proteolíticos. No se sabe con claridad qué es lo que ocurre a 37 grados, aunque estudios recientes indican que no es necesaria la penetración del interferón en la célula: conserva su actividad aun cuando esté unido a un soporte sólido, que deba mantenerlo fuera de la célula. Ruddle y otros investigadores han sugerido que la síntesis del receptor en las células humanas está controlada por el cromosoma número 21, porque sólo las células híbridas de ratón y humanas que contienen el cromosoma número 21 responden al interferón humano, y la respuesta es más intensa en las células que tienen una copia suplementaria del cromosoma.

El interferón mismo no es el agente antivírico. Su acción puede ser inhibida tratando las células con actinomicina D y cicloheximida; lo cual sugiere que actúa desencadenando un proceso consistente en la síntesis de ARN sobre el molde de ADN y en la síntesis de proteínas. Además, para que el interferón actúe contra los virus debe estar presente el núcleo celular. La interpretación más obvia es que el interferón cuyo lugar de acción estaría en la membrana celular, induce la síntesis de un nuevo ARN mensajero celular que es traducido, a su vez, dando lugar a otra proteína que sería la que realmente posee el efecto antivírico. La traducción de un mensajero de este tipo puede producir un gran número de moléculas proteicas antivíricas, de modo que un esquema tal constituiría un mecanismo de amplificación: así, una sola molécula de interferón podría desencadenar la síntesis de una gran cantidad de proteína antivírica. Esta hipótesis explicaría la intensísima actividad biológica del interferón; al parecer, basta con unas pocas moléculas del mismo para proteger a una célula.

La forma en que la proteína antivírica inhibe realmente la multiplicación de los virus es una cuestión que se ha estudiado en células infectadas, tratadas con interferón, y en sistemas acelulares que contenían componentes de las células tratadas. Los experimentos con células han puesto de manifiesto que el tratamiento con interferón no impide la pe-



PRODUCCION EN GRAN ESCALA de interferón de leucocitos humanos. Se centrifuga sangre y se recoge la fina capa correspondiente a los leucocitos (glóbulos blancos); los leucocitos están "contaminados" por algunos glóbulos rojos que deben eliminarse. Los glóbulos blancos purificados se suspenden en un medio de cultivo que contiene algo de suero sanguíneo humano o animal. Entonces se añade el

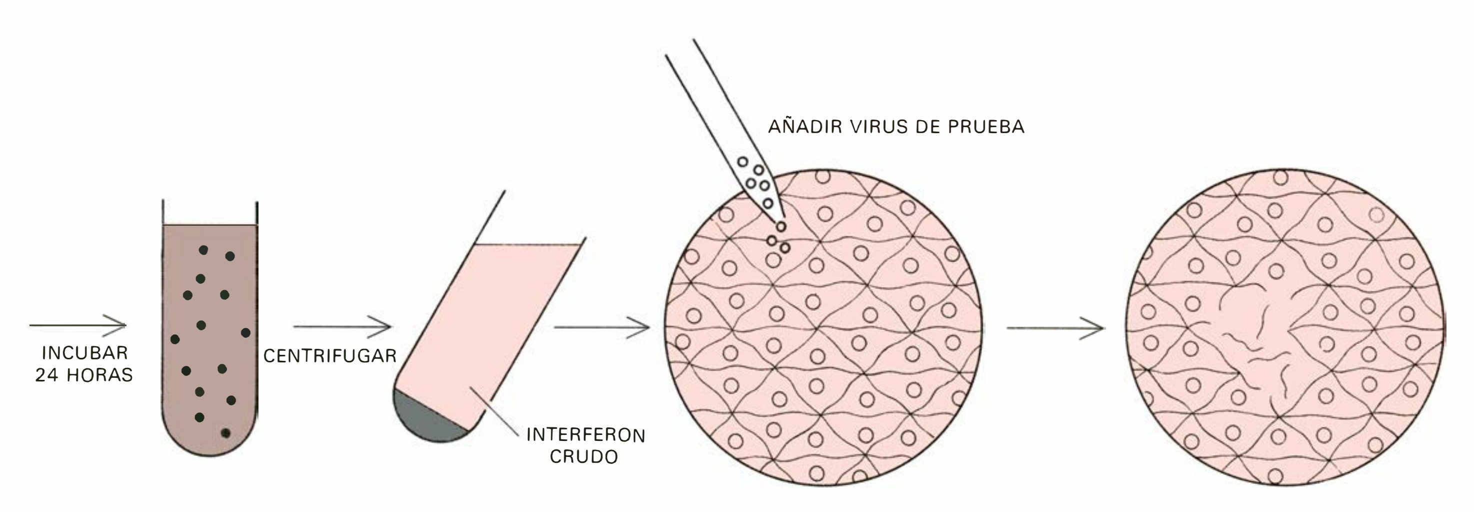
netración de las partículas víricas dentro de la célula. Hay pruebas claras de que lo que hace es inhibir la síntesis de ácido nucleico vírico o de proteínas víricas, o ambas síntesis; según las diferentes combinaciones de virus y célula se inhibe una síntesis o la otra, o ambas. El paso inhibido parece ser la síntesis de proteínas víricas en células infectadas por el virus vaccinia (vacuna) o por alguno de los pequeños virus de ARN, pero también se tienen pruebas claras de un efecto sobre la síntesis de ARN vírico en células infectadas por el pequeño virus tumorígeno SV40, que contiene ADN.

En cuanto a los sistemas libres de células, preparados a partir de células tratadas con interferón, parecen quedar afectados de diversas maneras, pero todas ellas relacionadas con la formación de nuevas proteínas. La iniciación de la síntesis de proteínas queda retardada. quizás en parte porque el tratamiento con interferón inhibe la modificación necesaria, a saber, la fijación de los ribosomas sobre la cabeza naciente de la extremidad de la molécula de ARN mensajero por la que empieza la traducción. Además, los extractos de células tratadas con interferón son más sensibles a los efectos del ARN de doble cadena, que es un inhibidor conocido de la iniciación de la síntesis de proteínas. La velocidad de crecimiento de la nueva molécula proteica también queda disminuida por el tratamiento con interferón. Quizás el tratamiento agote el suministro de determinado ARN de transferencia, necesario para la síntesis de proteínas, o bien induzca la formación de ribonucleasas que degraden los nuevos ARN mensajeros. Es evidente que el efecto del tratamiento con interferón sobre la síntesis proteica es complejo, lo cual sugiere que el interferón actúa como una hormona, que puede tener múltiples efectos, mejor que como un inhibidor metabólico con un efecto único sobre una etapa precisa de una vía biosintética particular.

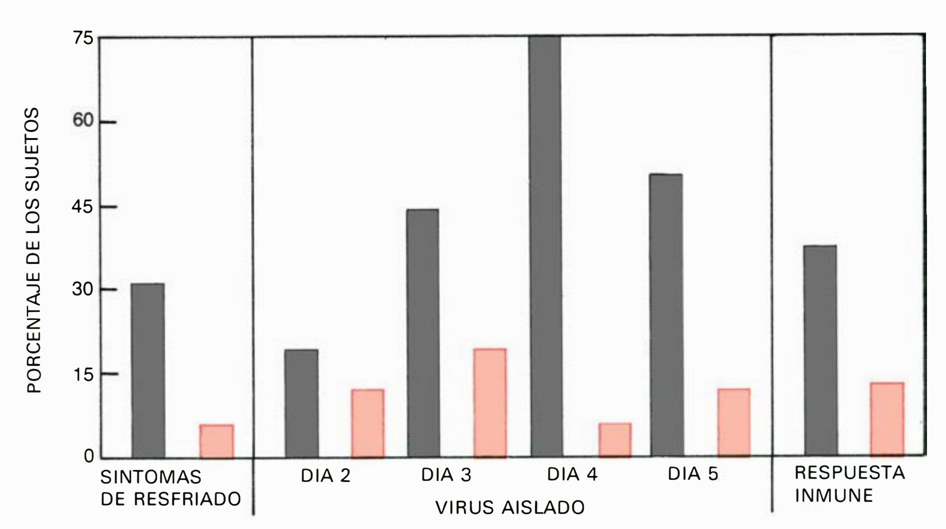
Si el interferón puede afectar a la síntesis de ácido nucleico y a la síntesis de proteínas, ¿por qué no mata a las células, a la vez que impide la multiplicación de los virus? Existen varias posibilidades. En primer lugar, la multiplicación vírica es un proceso exponencial, de forma que un ligero efecto sobre una etapa precoz puede verse aumentado en las fases siguientes. En segundo lugar, quizás el interferón inhiba tanto los procesos celulares como los procesos víricos, pero actúe sólo en células que están infectadas por virus; en tal caso morirían las células infectadas, pero la gran mayoría de las células no quedaría afectada. En tercer lugar, hay indicios de que los ARN mensajeros víricos tienen una mayor avidez por los componentes de la maquinaria celular de síntesis proteica que los ARN mensajeros celulares y quizás aquellos queden más fuertemente afectados por el tratamiento con interferón que las propias células. En cualquier caso, la discriminación del interferón entre virus y células no es absoluta: el interferón retarda el crecimiento de las células, pero a unas concentraciones mucho mayores que las requeridas para bloquear el crecimiento de los virus. Se han estudiado otros efectos del interferón. Así se sabe que refuerza varios mecanismos de defensa celulares distintos, por mediación de diferentes tipos de células, y hace variar la forma de expresión de algunos antígenos de la superficie celular. El interferón se nos presenta de una forma cada vez más clara como un agente versátil que posee efectos sobre virus y sobre células huésped, si bien su efecto sobre los virus es mayor.

Ya en 1960, un proyecto de investigación conjunto organizado en Gran Bretaña por el Medical Research Council y tres compañías de productos farmacéuticos (Imperial Chemical Industries, Glaxo Laboratories y la Wellcome Foundation) adoptó como objetivo la determinación del valor del interferón como agente antivírico en humanos. El problema inicial, que sigue vigente, fue preparar una cantidad suficiente de interferón que fuese activo en células humanas e inofensivo al administrarlo a sujetos humanos. (El primer ensayo se hizo con interferón de mono, pero resultó excesivamente caro; las pruebas siguientes se hicieron con interferón humano.) Otro problema era idear situaciones de prueba en las que se pudiera administrar de forma eficaz una cantidad adecuada de interferón y se pudiera observar una respuesta, que fuera medible y controlada, a la infección vírica. El grupo de Norman B. Finter, de Imperial Chemical Industries, realizó experimentos en ratones que proporcionaron la información básica para la planificación de pruebas clínicas.

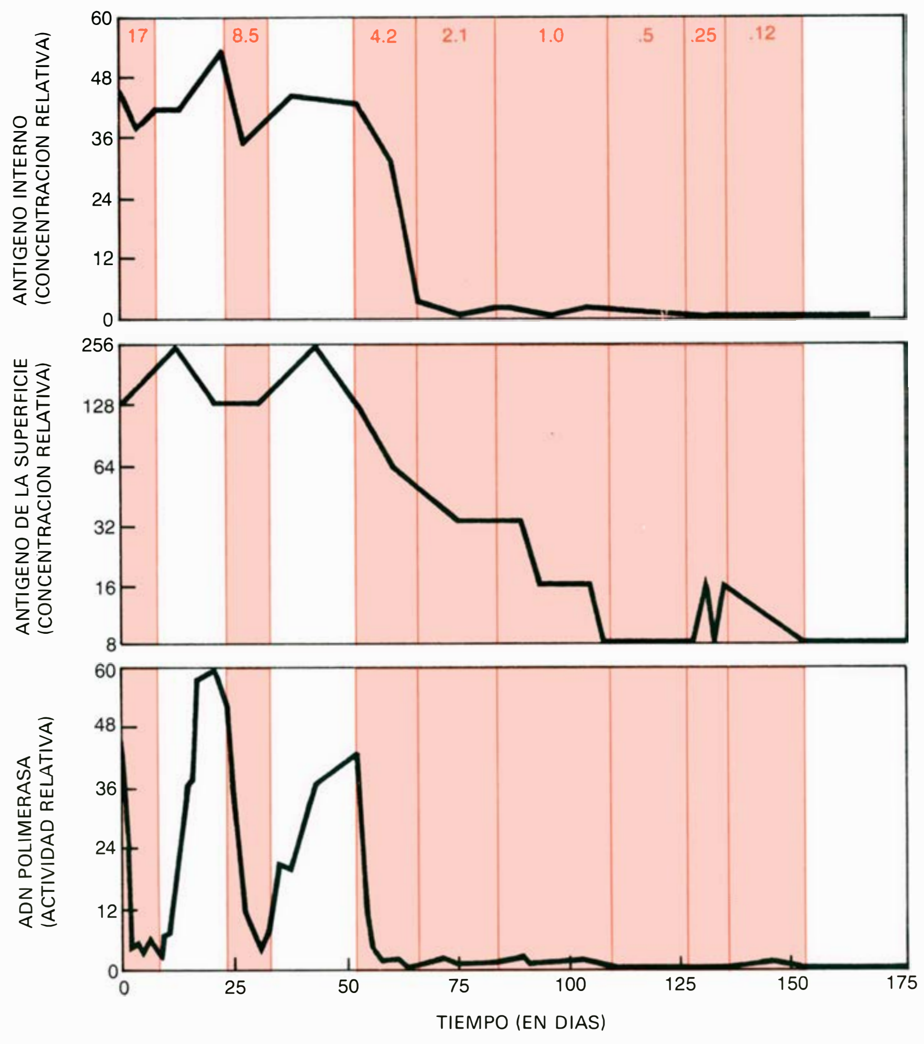
La primera prueba, realizada en 1961, consistió en inyectar interferón de mono en determinado punto del brazo de voluntarios que nunca habían sido vacunados contra la viruela y en otro punto diferente inyectarles un placebo. Al día



inductor, que es el virus de Sendai; se incuba durante 24 horas la suspensión de células con el virus, período durante el cual las células producen interferón ***(en color).*** **Se centrifuga la suspensión para eliminar las células y los virus; el sobrenadante es un extracto crudo de interferón. Se comprueba su autenticidad determinando su capacidad para reducir las zonas de células muertas en cultivo expuesto al ataque vírico.**



EL VIRUS DEL RESFRIADO COMUN (rhinovirus) se atomizó en el interior de la nariz de voluntarios, la mitad de los cuales fueron previamente tratados con interferón, también por atomización nasal (el tratamiento duró un día antes de la exposición al virus y tres días después). Los sujetos tratados desarrollaron menos síntomas de resfriado común *(franjas de color)* que los sujetos control no tratados *(franjas grises)*. En los días indicados, era menor la cantidad de virus que se podía aislar de la nariz de los sujetos tratados, y produjeron menos anticuerpo del virus.



LA INFECCION CRONICA por hepatitis *B* se trató con interferón en una prueba realizada por Harry B. Greenberg y sus colaboradores, de la Stanford University School of Medicine. En el caso que aquí se ilustra, se inyectó interferón durante tres períodos distintos; las cifras superiores indican la dosis en múltiplos de 10 000 unidades por kilogramo de peso corporal y por día. Las curvas representan el efecto del interferón sobre los niveles respectivos, en la sangre del paciente, de dos antígenos y de un enzima, la ADN polimerasa, que están asociados con las "partículas Dane" presentes en la sangre de los pacientes de hepatitis *B* y que, según se cree, son el virus de la hepatitis *B*. Los períodos cortos de grandes dosis de interferón tuvieron un efecto claro, aunque pasajero; un período continuado de dosis pequeñas pareció tener como resultado la supresión prolongada de la actividad del supuesto virus. (Ilustración realizada por Jerome Kuhl.)

siguiente, ambos puntos se exponían al virus vaccinia. Se desarrollaron lesiones debidas al virus en 37 de los 38 puntos control y sólo en 14 de los 38 puntos previamente tratados con interferón, lo que constituía una diferencia altamente significativa. Pero el planteamiento de esta prueba favorecía deliberadamente un resultado positivo: el interferón se inyectaba localmente en un número limitado de células y estas mismas células eran las que se exponían al virus. Se logró un éxito similar en los resultados de una prueba del mismo tipo que medía los efectos del interferón sobre el virus vaccinia en el ojo de conejos.

Fue mucho más difícil demostrar una protección contra infecciones menos localizadas. Resultaron negativas dos pruebas, realizadas en 1965 y 1970, en que se administraba interferón a sujetos con infecciones víricas respiratorias. Analizando ahora el fracaso se deduce que no se les dio una cantidad de interferón suficiente como para proteger las células de las vías respiratorias superiores, que son muy eficaces a la hora de eliminar el material extraño –incluso cuando es un agente potencialmente protector– por el movimiento ondulante de sus cilios. No obstante, este resultado desalentador sirvió para desacreditar el enfoque del interferón exógeno. Durante algún tiempo, pasamos a estudiar el otro enfoque posible: la administración de poli *rI:rC* y otros nucleótidos, como drogas antivíricas, con la esperanza de inducir la producción de interferón endógeno en las propias células del voluntario. Al final de la década de los años sesenta, otros varios laboratorios enfocaron también el problema de la misma forma. A pesar de que los ARN de doble cadena estimulaban la producción de grandes cantidades de interferón circulante en ratones y otros roedores experimentales, la cantidad de proteína elaborada en monos y en seres humanos era mucho menor, sin duda porque dichos ARN eran destruidos por un enzima del suero. Tendían a provocar fiebres altas y producían asimismo otros efectos tóxicos en los voluntarios. La conclusión fue que los nucleótidos de doble cadena, al menos en su forma no modificada, no resultarían ser agentes clínicos eficaces e inofensivos en el hombre. Recientemente, sin embargo, Hilton B. Levy y sus colaboradores de los National Institutes of Health y el grupo de Carter de Roswell Park han podido incrementar la relación actividad-toxicidad de algunos ARN de doble cadena por modificación química de su molécula, y hay programadas unas pruebas

clínicas con estos agentes modificados.

Cuando parecía que los nucleótidos de doble cadena no servirían como inductores de interferón endógeno, el interés de los investigadores en clínica se volvió a centrar en la proteína exógena. Kari Cantell, del Instituto Nacional Serológico, de Helsinki, desarrolló un método para la producción y purificación de interferón humano en gran escala, a partir de leucocitos conseguidos en los bancos de sangre. El interferón elaborado por el grupo de Cantell ha permitido la realización de un gran número de pruebas clínicas.

En 1972, un grupo de la unidad de estudio del resfriado común del Medical Research Council de Salisbury puso a prueba el interferón frente a infecciones por virus de gripe y de resfriado común. Una dosis total de 800 000 unidades de interferón humano por voluntario sólo lograba retrasar ligeramente el inicio de la infección por el virus de la gripe. Cantidades muy superiores (14 millones de unidades por paciente, administradas en 39 dosis durante un período de cuatro días) sí eran capaces de dar como resultado un descenso, estadísticamente significativo, de los síntomas de infección por resfriado común y de la presencia de virus en el material extraído de la nariz. Este fue el primer ensayo positivo del interferón sobre la infección del resfriado común; de particular significación fue el hecho de que no se observaran efectos tóxicos. Sin embargo, no resulta práctico administrar interferón para prevenir los resfriados, aunque sólo sea porque el coste actual de un millón de unidades de interferón es aproximadamente 50 dólares.

Sin embargo, hay infecciones víricas que pueden amenazar la vida de algunas personas. Los pacientes que tienen un trastorno inmunológico crónico o cuyo sistema inmune ha sido deprimido por drogas administradas para impedir el rechazo de un órgano trasplantado o para tratar el cáncer, son desacostumbradamente susceptibles a una amplia gama de infecciones víricas. Thomas C. Merigan, de la Stanford University School of Medicine, ha comprobado que una dosis diaria de interferón variable entre 2,5 y 10 millones de unidades puede disminuir tanto la propagación como el dolor de la infección por varicela-zoster en pacientes afectados por tumor maligno; en otra prueba, todavía en curso, se utiliza 20 millones de unidades diarias. Merigan ha informado que, al parecer, se necesita una dosis igualmente elevada para reducir la excreción de virus en los niños recién nacidos (cuyo sistema inmune aún no está totalmente desarrollado) que presentan una infección crónica por citomegalovirus.

El resultado clínico más significativo obtenido hasta el momento quizá sea el logrado por un grupo encabezado por Merigan y William S. Robinson, en Stanford. Observaron el sorprendente efecto del interferón sobre una infección crónica asociada con la hepatitis sérica, o hepatitis *B*. Aproximadamente el 10 por ciento de los pacientes con hepatitis *B* hospitalizados en los Estados Unidos sufren una infección crónica por este virus. Se ha estimado que, en todo el mundo, hay unos 100 millones de personas que padecen esta condición crónica; están expuestos a ataques repetidos de hepatitis activa, con las consiguientes lesiones hepáticas; pueden, además, contagiar la enfermedad a otras personas. El tratamiento con dosis de interferón comprendidas entre 500 000 y 10 millones de unidades diarias provocó un descenso rápido del nivel de varios indicadores sanguíneos de la hepatitis: dos antígenos y un enzima asociados con las partículas que, según se cree, son las del virus responsable de la enfermedad. El efecto era pasajero cuando el interferón se administraba durante 10 días o menos, pero persistía mientras se continuaba dando pequeñas dosis e incluso durante algún tiempo posterior al tratamiento. En otras palabras, parece que el interferón interfiere la multiplicación de las supuestas partículas víricas. Ahora se está ampliando la prueba al estudio del efecto del interferón en los períodos de hepatitis activa.

Hans Strander y sus colaboradores, del Instituto Karolinska de Estocolmo, están llevando a cabo una prueba interesantísima relacionada con los efectos del interferón en el sarcoma osteogénico, enfermedad cancerosa poco común que quizá sea iniciada por un virus. Es una enfermedad grave, en la que tan sólo un 20 por ciento de los enfermos consiguen sobrevivir dos años. En esta prueba, se operaba primero a los pacientes y, en algunos casos, se les sometía a una radioterapia para reducir el volumen del tumor. Luego se les inyectaba entre dos y tres millones de unidades de interferón, tres veces por semana durante 18 meses. Hasta ahora, se ha tratado a veintiún pacientes, sin que se haya observado efectos secundarios producidos por el interferón. En comparación con otros casos similares registrados antes, parece hacer un descenso en la frecuencia de la metástasis (propagación del cáncer a los demás tejidos), que suele ser la causa de la muerte. Los resultados son esperanzadores y la prueba sigue. Si el interferón influye realmente en el curso de esta enfermedad, es difícil saber cómo actúa. Quizás ejerza su acción sobre algún virus no identificado que tenga que ver con la propagación del cáncer. O quizá funcione de un modo más general, como agente de control celular que retarde el crecimiento de las células cancerosas: la pérdida del control del crecimiento y la proliferación es una de las características de las células cancerosas. Martin S. Hirsch y sus colaboradores, del Massachusetts General Hospital, han empezado ahora a estudiar el efecto del interferón sobre varios tipos de cáncer de los tejidos linfáticos en pacientes a los que se les ha trasplantado un riñón.

Los resultados de las pruebas citadas reclaman claramente la continuación de las investigaciones clínicas en relación con el efecto del interferón sobre las enfermedades cancerosas y sobre las infecciones víricas. Dichas pruebas precisan grandes cantidades de interferón (los National Institutes of Health norteamericanos acaban de aprobar un presupuesto de 1 millón de dólares para la fabricación de interferones humanos y de ratón). Sin embargo, a largo plazo, se necesita un interferón más barato, y se está investigando varias maneras de alcanzar este objetivo. Una de las posibilidades es hacer crecer un gran número de linfoblastos humanos, que son los precursores de las células responsables de la inmunidad, e inducir luego a estas células a fabricar interferón; otra posibilidad es alterar alguna línea de células humanas por manipulación genética a fin de que produzcan una mayor cantidad de interferón. Quizás algún día sea posible, a través de la manipulación genética, introducir en las bacterias los genes que codifican la producción de interferón, pudiéndose entonces utilizar, como fábricas de interferón, a extensos cultivos de bacterias.

Hasta el momento, se ha establecido con claridad que el interferón es un inhibidor específico y potentísimo de la multiplicación vírica. Hay bastantes indicios de que, en realidad, es una sustancia reguladora de varias funciones celulares y que su actividad antivírica es tan sólo una de sus propiedades. Cada vez tenemos un mayor número de razones para pensar que podría ser un valioso agente en medicina clínica; no para proteger a toda la población de los resfriados y de la gripe, sino para proteger a los pacientes para quienes las infecciones víricas son letales en potencia.

Ciencia y sociedad

Corrección del dogma

El desciframiento de la biología molecular del gen ha originado varios "dogmas". El más básico, tal vez, está expresado en la frase "un gen, una proteína"; significa que la información necesaria para la síntesis de cada proteína está codificada en un gen único para esa proteína; conociendo la secuencia de codones (grupos de tres unidades codificadoras) que representa un gen en la molécula lineal de ADN que detenta el plan genético del organismo, se puede deducir la secuencia singular de aminoácidos de la proteína especificada por ese gen. Aunque no se ha abolido el dogma, sí ha sufrido una corrección fundamental por los recientes trabajos de Frederick Sanger y sus colegas

Durante algún tiempo, se han venido acumulando pruebas respecto de que la información contenida en el ADN de uno de los virus bacterianos más pequeños, denominado ΦX174, era insuficiente para codificar el proceso necesario para los nueve tipos de moléculas proteicas que el virus podía fabricar al invadir una célula huésped. Aunque no se han establecido con exactitud los papeles desempeñados por las diferentes proteínas, algunas participan en la duplicación del ADN del virus y otras actúan proporcionando el cápside, o cubierta en la cual viene plegado apretadamente el ADN. El virus tiene una estructura icosaédrica, de unos 25 nanómetros de diámetro. Al compararse los cálculos estimativos sobre el número total de aminoácidos presentes en las nueve proteínas producidas con el número que podría estar codificado por la molécula de una cadena de ADN del virus, se hizo evidente que a la molécula de ADN le faltaba entre un 10 y un 15 por ciento de longitud. Se sabía que el ADN constaba de unos 5400 nucleótidos. Para especificar cada aminoácido se precisa una secuencia de tres nucleótidos, es decir, un codón. Los codones vienen designados por tripletes que combinan las cuatro letras *A*, *T*, *G* y *C*, que significan, respectivamente, adenina, timina, guanina y citosina, y son los substituyentes que distinguen a los cuatro nucleótidos. En total hay 64 codones, que representan todas las combinaciones posibles de las cuatro letras tomadas de tres en tres. Cada codón (exceptuando tres codones "sin sentido") especifica uno de los veinte aminoácidos que sirven como bloques para construir proteínas. Puesto que hay 61 codones "con sentido", un aminoácido determinado viene normalmente definido por más de un codón. Así, el aminoácido valina puede ser especificado por cualquiera de los cuatro codones siguientes: *GTA*, *GTC*, *GTG* o *GTT*. Por otra parte, la metionina está representada por un solo codón, *ATG*, que también sirve como iniciador, dando la señal para dar comienzo a una cadena proteínica. Se obtuvieron pruebas concluyentes de que el ADN del ΦX174, si se traduce de la manera normal, consta de un número demasiado pequeño de nucleótidos como para responder por las nueve proteínas del virus; ello ocurrió el otoño pasado cuando Sanger y sus colaboradores publicaron la secuencia completa de nucleótidos del fago ΦX174 y esclarecieron simultáneamente el misterio de la información genética que faltaba. Cuando compararon las secuencias de nucleóticos del ADN y las secuencias de aminoácidos de las proteínas correspondientes, hallaron que, en dos casos, el código para una segunda proteína venía encajado en la secuencia de nucleótidos para una proteína de mayor tamaño. Así pues, el código para el producto *B* del gen estaba encajado en el código para el producto *D* del gen. En otras palabras, con dos genes que normalmente originarían sólo dos proteínas, el ΦX174 podía fabricar cuatro proteínas claramente diferentes.

Además, en cada caso, la proteína más pequeña no es simplemente una porción truncada de otra mayor; representa una lectura completamente diferente de tripletes de codón lograda mediante un desplazamiento de fase. Por ejemplo, la secuencia de nucleótidos que se analiza como... *G*, *AAG*, *TTA*, *ACA*... para especificar a los aminoácidos lisina, leucina y treonina de la proteína *A*, se lee desde un punto situado un nucleótido más atrás en la síntesis de la proteína *B*, produciendo los codones... *GAA*, *GTT*, *AAC*, *A*..., que especifican a los aminoácidos glutamina, valina y asparagina. Hay que destacar que, de los 1536 nucleótidos (513 codones) del gen *A*, puede recortarse un tramo de 360 nucleótidos (120 codones) y que, con un desplazamiento de fase de un nucleótido, da origen a una proteína viable *B*, totalmente distinta de cualquier parte de la proteína *A*.

Al parecer, nadie había imaginado la posibilidad de semejantes lecturas con desplazamiento de fase hasta que el grupo de Sanger descubrió su existencia. Considerándolo retrospectivamente, había varios motivos válidos para pasar por alto esa posibilidad. En casi todos los organismos en los que se puede comparar la longitud del ADN y la longitud total de la proteína producida, suele haber un exceso manifiesto de ADN. Lo que ha intrigado a los investigadores no es la falta de ADN sino, más bien, la presencia de tanto ADN "silencioso". La excepción presentada por el φX174 no parecía grave

Lo que también obraba contra la posibilidad de una doble lectura era la existencia de los tres codones "sin sentido", *TAG*, *TAA* y *TGA* que, a menudo, actúan como finalizadores de cadena por carecer de aminoácidos que sean su réplica. Por consiguiente, en cualquier tramo completamente aleatorio de 64 codones cabría hallar tres codones sin sentido. En la lectura con desplazamiento de fase de los genes *A* y *B* del fago ΦX174 no hay codones sin sentido en 633 tripletes solapados. Aun cuando alguien hubiese imaginado semejante posibilidad, parecería aún más improbable que una sola lectura desplazada de fase diera origen a una segunda proteína viable, siendo que una sola mutación en un gen (el cambio de un nucleótido entre varios centenares) da comúnmente origen a una proteína que no puede funcionar.

A pesar de la corrección de la hipótesis de "un gen, una proteína", no se cree que la producción de más de una proteína por parte de un gen sea un suceso corriente. Se sospecha que puede haberse suscitado en el caso del fago ΦX174, y quizá en virus pequeños parecidos, porque su cubierta proteica establece un límite absoluto sobre el contenido de ADN del virus. Para que el virus desarrolle cualquier función nueva en el curso de la evolución, podrá hacerlo sólo si se extrae información adicional de una cantidad fija de ADN.

La física del balanceo

Cualquiera sabe cómo poner en movimiento un columpio de jardín balanceándolo, pero pocos están familiarizados con los principios físicos que ello supone. En *The American Journal of Physics*, Stephen M. Curry de la Universidad de Texas, en Dallas, analiza la mecánica del columpiarse y lo compara con el funcionamiento de un amplificador paramétrico, es decir, con un sistema oscilante en el cual la ganancia se produce variando un solo parámetro, por ejemplo, lo constante elástica de un muelle, o la capacitancia de un circuito eléctrico.

Curry inicia su análisis con un niño que balancea un columpio mediante el método de ponerse de pies y en cuclillas, alternativamente. Este movimiento hacia arriba y hacia abajo cambia periódicamente la longitud del columpio, o la distancia del centro de giro al centro de masa. Debido a que, incluso en reposo, el columpio no está perfectamente vertical, sino que posee cierta leve energía inicial, el balanceo del niño aumenta su amplitud en un leve incremento por cada ciclo de movimiento. Conforme va creciendo la amplitud del columpio, aumenta la fuerza centrífuga y el niño tiene que aplicar una mayor cantidad de fuerza en cada oscilación. La suma de la energía cinética y potencial generada por su trabajo mecánico se compone por tanto periódicamente, como ocurre con el interés bancario, y aumenta exponencialmente en relación con el tiempo.

Contrariamente a lo que sugiere la intuición, el peso no es una ventaja para columpiarse. Aunque un niño delgado realiza menos trabajo que otro obeso contra las fuerzas gravitatorias y centrífugas, para ponerlo a columpiarse se precisa proporcionalmente menos energía. En cambio, la altura es una clara ventaja: cuanto mayor sea la distancia a lo largo de la cual pueda el niño desplazar su centro de masa, mayor será el cambio de amplitud en cada oscilación y mayor también el ritmo de aumento de energía.

En teoría, no se puede balancear un columpio desde la posición de reposo, si su energía inicial es cero. Esta paradoja aparente puede entenderse intuitivamente basándose en la simetría: si el columpio está fijo en la posición vertical, nada hay que indique ninguna dirección específica de movimiento lateral, y las componentes ascendente y descendente del movimiento del niño se anulan entre sí. Sin embargo, en la práctica, no puede darse esa condición de energía inicial nula. Aun cuando descendiera la temperatura hasta el cero absoluto, el columpio seguirá teniendo una pequeñísima energía a causa de la imposibilidad, según la necánica cuántica, de conocer exactamente la posición y el momento del centro de masa. Curry demuestra que en este caso, el peor posible, la pequeñísima energía inicial del columpio puede incrementarse hasta un nivel macroscópico en un tiempo sorprendentemente breve. Para un típico columpio de tres metros, y con un cambio periódico de longitud de 40 centímetros, la energía almacenada alcanzará hasta varios centenares de julios (un factor de incremento de 10^{36}), al cabo de unos seis minutos.

La mecánica de la acción de columpiarse mientras se está sentado es más compleja y exige un tipo diferente de análisis. Para balancearse, el niño se inclina súbitamente hacia atrás, de modo que su cuerpo adquiera un momento angular neto en torno a su centro de masa, desplazándolo desde su posición inicial. Este par aplicado periódicamente inicia el movimiento de balanceo.

A diferencia del balanceo en pie, en el cual la potencia de salida aumenta con la amplitud, el par aplicado en la posición de sentado permanece aproximadamente constante, y la energía total del sistema se incrementa linealmente y no exponencialmente con el tiempo. Concluye Curry que la mejor estrategia para columpiarse a partir de la posición de reposo consiste en balancearse en posición de sentado hasta que se alcance cierta energía óptima y luego recurrir a la posición erecta para aumentar más la amplitud del movimiento.

Del bronce al acero

Cómo terminó la Edad de Bronce y comenzó la Edad de Hierro? Una hipótesis ampliamente aceptada conecta esa transición con los confusos desórdenes ocurridos en el Mediterráneo oriental, que se iniciaron alrededor del 1200 a.C., y que incluyeron, entre otros sucesos violentos, la destrucción de las ciudades-estado micénicas de Grecia realizada por misteriosos invasores conocidos como "pueblos del mar". La hipótesis continúa afirmando que, al estar privados de acceso al estaño, y quizás también al cobre, los herreros de ese período se vieron obligados a usar más el hierro, metal conocido y abundante, que anteriormente se había fundido para fines más ornamentales que utilitarios. Trabajando con hierro por no tener alternativa, los herreros descubrieron pronto sus numerosas ventajas sobre el bronce. La consecuencia de esto fue que, si bien los ingredientes para el bronce volvieron a estar disponibles alrededor del año 900 a.C., el metal más antiguo jamás recobró su anterior posición.

Esta interpretación de los hechos no puede ser correcta, argumenta Robert Maddin, profesor de metalurgia de la Universidad de Pennsylvania. En primer lugar, el bronce es superior al hierro como material para herramientas y armas. El hierro se deforma al ser sometido a esfuerzos de 1406 a 2109 kilogramos por centímetro cuadrado, mientras que esto no le ocurre al bronce sino hasta que el esfuerzo fluctúa entre 4220 y 4922 kg/cm². Se puede aumentar la resistencia de ambos metales mediante forja en frío; este proceso aumenta la tenacidad del hierro hasta unos 7030 kg/cm² y la del bronce hasta unos 8437 kg/cm². El hierro no se funde, por tanto no puede moldearse, a temperaturas por debajo de los 1500° Celsius (un nivel que los primitivos herreros no lograban fácilmente) mientras que el bronce se funde a poco más de 1000° Celsius, o sea, la temperatura usada para cocer cerámica. Por último, el hierro se corroe rápidamente y el bronce se oxida con mucha lentitud.

¿Por qué, entonces, eclipsó el hierro al bronce súbitamente en las postrimerías del segundo milenio antes de Cristo? Porque, dice Maddin, realmente no era hierro lo que se usaba, sino acero. En un reciente simposio sobre análisis de laboratorio en arqueología, celebrado en la Universidad de Pennsylvania, Maddin comunicó los resultados de estudios metalúrgicos de herramientas y armas de los inicios de la Edad de Hierro, procedentes de excavaciones hechas en Israel. Con la cooperación de funcionarios de museos, cortó por la mitad varios artefactos de hierro corroído, dejando así al descubierto metal sin corroer. El examen microscópico reveló que el metal interior contenía de tres a cinco por mil de carbón. Maddin calculaba que el metal dispuesto en la superficie de esos objetos habría contenido alrededor del uno por ciento de carbón. Un hierro que haya sido "acerado" con tanto carbón no se deformará bajo tensiones de menos de 9843 kilogramos por centímetro cuadrado; el forjado aumenta la tenacidad del metal hasta 14.061 kilogramos por centímetro cuadrado.

Maddin sugiere que los primitivos productores de acero al carbono no intentaron conscientemente cementar el hierro, pero pueden haber creído que un calentamiento prolongado e intenso en las hogueras de carbón del herrero purificaba de algún modo el metal. El hierro absorbe carbón a una velocidad que se relaciona con la temperatura del fuego; la cementación a 1100° Celsius es más de dos veces más rápida que la que se produce a 1000 grados. El templar el metal incandescente en agua, proceso que los antiguos herreros presumiblemente inventaron por su propia seguridad cuando comenzaron a trabajar en hierro (un artefacto de bronce fundido podía dejarse en su molde hasta que se enfriaba), endurece el acero al carbono más que la forja, aunque el proceso hace frágil al metal. En cualquier caso, aquellos que usaron hachas, azuelas, picos, palas, espadas, puñales, puntas de lanza y de flecha, de acero, quedaron en lo sucesivo satisfechos con la idea de dejar que el bronce cumpliera la función ornamental que antes desempeñaba el hierro.

Algoritmos

Un algoritmo es un sistema de reglas que permiten obtener una salida específica a partir de una entrada específica. Cada paso debe estar definido exactamente, de forma que pueda traducirse a lenguaje de computador

Donald E. Knuth

Hace diez años, a la mayoría de las personas cultas, el término "algoritmo" les era desconocido; en realidad, apenas si resultaba necesario el vocablo. El rápido desarrollo de las ciencias de la computación, enfocadas en el estudio de algoritmos, ha cambiado la situación: el término se hace ahora imprescindible. Existen algunas otras palabras que casi, aunque no del todo, captan la noción requerida: procedimiento, receta, proceso, rutina, carrerilla. Como también designan esas palabras, un algoritmo es un conjunto de reglas o instrucciones que permiten obtener una salida específica a partir de una entrada específica. El rasgo característico de los algoritmos es que, en ellos, toda vaguedad o imprecisión debe eliminarse: sus reglas tienen que describir operaciones simples y bien definidas, de suerte que un máquina pueda ejecutarlas. Además, todo algoritmo ha de concluir en un número finito de pasos.

Un programa es la expresión de un algoritmo, formulado en un lenguaje bien definido. Así pues, un programa de computador representa un algoritmo, aunque, en sí mismos, los algoritmos son concepciones mentales y existen independientemente de sus representaciones, tal como la noción "número 2" existe en nuestras mentes sin necesidad de escribir dicha cifra. Quien haya puesto a punto un programa de computador habrá de convenir en que un algoritmo debe estar definido con gran precisión, y que su preparación exige prestar a todos los detalles una atención minuciosa, insólita en comparación con las restantes actividades humanas.

Hacia 1800 a.C., matemáticos babilónicos de la época de Hammurabi dieron reglas de solución de muchos tipos de ecuaciones y escribieron programas para resolver problemas numéricos. Estas reglas consistían en procedimientos a seguir paso a paso, aplicados sistemáticamente a casos numéricos particulares. El término "algoritmo" se originó también en el Oriente Medio, aunque mucho más tarde. Procede del último nombre del erudito persa Abu Ja'far Mohamed ibn Mûsâ al-Khowârizmî, cuyo tratado de aritmética (hacia 825 d. C.) tuvo gran influencia a lo largo de muchos siglos.

Tradicionalmente, los algoritmos sólo guardaban relación con el cálculo numérico. Pero la experiencia recogida con los computadores ha puesto de manifiesto que unos datos manipulados mediante un programa pueden representar prácticamente cualquier cosa. Consecuentemente, las ciencias de la computación han desplazado ahora su actividad hacia el estudio de diversas estructuras mediante las que representan información, así como a los aspectos de la ramificación, o adopción de decisiones, de los algoritmos, que les permitan seguir una u otra secuencia de operaciones, de acuerdo con la situación en ese instante. Son precisamente estas características de los algoritmos las que hacen que, en ocasiones, los modelos algorítmicos de organización y representación de conocimientos sean superiores a los modelos matemáticos tradicionales. A pesar de que los algoritmos numéricos tienen, sin duda alguna, importantes características, me ceñiré en la exposición subsiguiente al estudio de ejemplos no numéricos, a fin de subrayar que los algoritmos tratan del manejo de símbolos que pueden, o no, representar números.

Exploración de la memoria de un computador

Analizaré con cierto detalle un sencillo problema de recuperación de una información, al objeto de mostrar cómo los algoritmos pueden estudiarse provechosamente. El problema consiste en descubrir si una cierta palabra, x, figura o no en una tabla de palabras almacenadas en la memoria de un computador. La palabra x podría ser el nombre de una persona, el número de serie de una pieza, una palabra de un idioma extranjero, un compuesto químico, el número de una tarjeta de crédito, o cualquier otra cosa. El problema tan sólo ofrece interés cuando el conjunto de todas las x posibles sea tan grande que el computador no tenga acceso inmediato a todas ellas, pues, en este caso, bastaría reservar un registro de la memoria para cada palabra.

Supongamos que en la memoria del computador se hayan almacenado n palabras. El problema consiste en preparar un algoritmo que acepte como entrada la letra x, y cuya salida sea el lugar j en que se encuentra x. Así pues, si x se encuentra en la memoria, la salida será un número entre 1 y n; por otra parte, si x no está en ella, la salida deberá ser 0, para indicar que la búsqueda ha sido infructuosa.

Este problema se resuelve fácilmente. El algoritmo más sencillo consiste en almacenar las palabras en los registros, de 1 hasta n, e ir mirando cada palabra sucesivamente. Si x se encuentra en el j-ésimo registro, el computador debería mostrar j en su salida, y detenerse; pero si el computador agota infructuosamente las n posibilidades, debería mostrar 0 y detenerse.

Sin embargo, y con toda probabilidad, la estrategia de búsqueda así descrita no sea bastante precisa para un computador por lo que el procedimiento deberá enunciarse más rigurosamente. Puede expresarse como una sucesión de pasos, de la forma siguiente:

Algoritmo A; exploración secuencial.

A1. [Partida]. Poner $j \leftarrow n$. (La flecha indica que la variable j toma el valor n, número de palabras de la tabla a explorar. Este es el valor inicial de j. En etapas subsiguientes, el algoritmo hará que j recorra la sucesión de valores n, $(n-1)$, $(n-2)$, etc., hasta que se alcance el valor 0 o un registro que contenga la palabra x.)

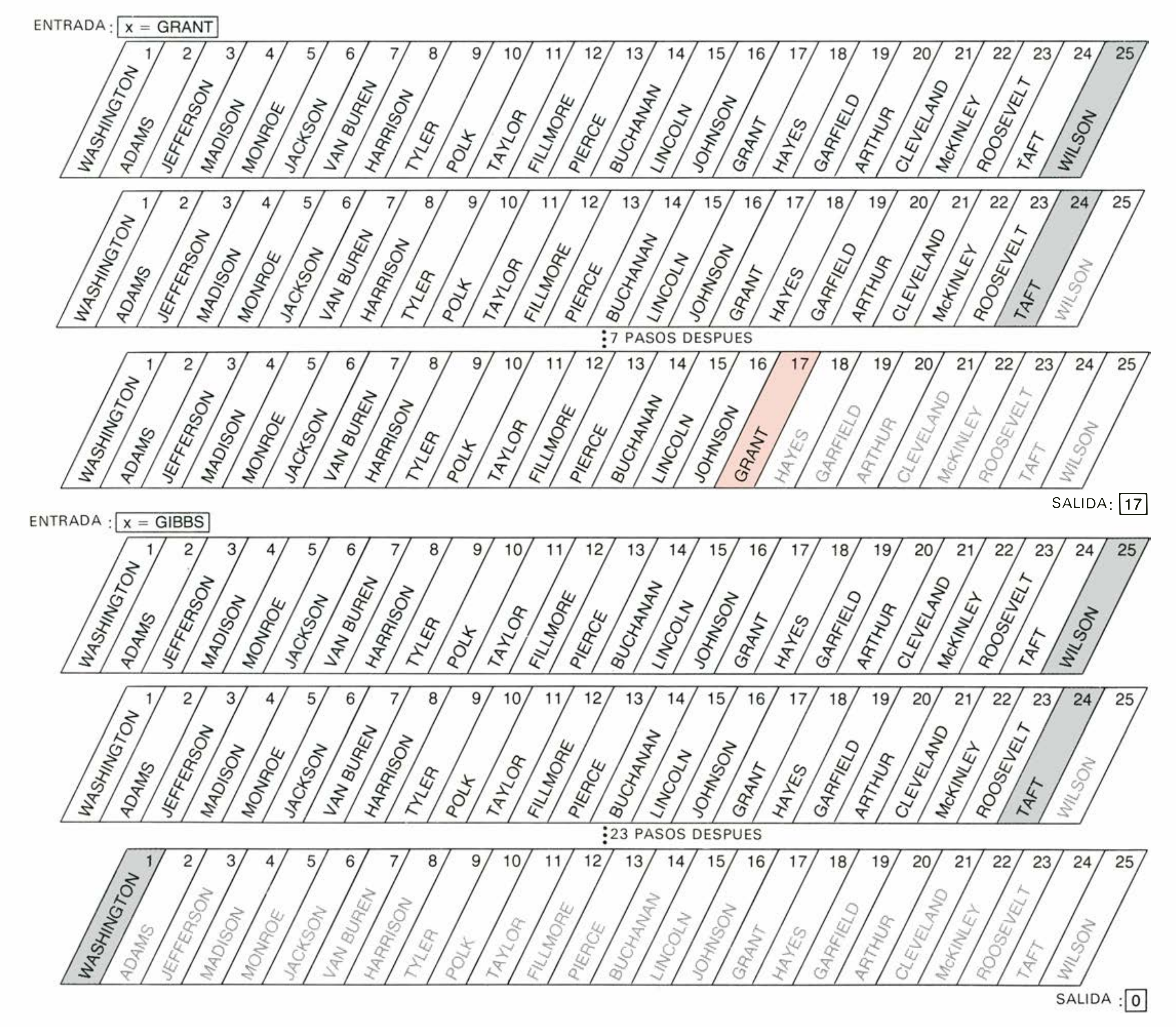
A2. [¿Fracaso?] Si $j=0$, dar salida 0 y finalizar el algoritmo. (En otro caso, proseguir al paso A3.)

A3. [¿Exito?] Si $x=$CLAVE $[j]$, dar salida j y finalizar el algoritmo. (El término CLAVE $[j]$ representa la palabra situada en el registro de lugar j.)

A4. [Repetición]. Poner $j \leftarrow j-1$ (disminuir en 1 el valor de j) y volver al paso A2.

Los pasos o etapas de este algoritmo pueden visualizarse por medio de un diagrama de circulación *(véase la ilustración de la página 44)*. La importancia de tener que especificar cuidadosamente cada paso se debe, entre otras razones, a que el algoritmo debe funcionar en todos los casos. Así, por ejemplo, la somera descripción inicial podría sugerir un algoritmo defectuoso, que saltaría directamente desde el paso A1 hasta el A3; tal algoritmo fallaría cuando $n=0$ (cuando la tabla no contuviera ninguna palabra), pues el paso A1 situaría j en 0, y el paso A3 buscaría en vano la inexistente CLAVE [0].

Vale la pena observar que el algoritmo A podría mejorarse dándole sentido al término CLAVE [0]. Para ello sería preciso admitir que una palabra puede ser alojada, además de en los registros 1 a n, en el "registro 0". En este caso, si el paso A1 enunciase "llevar CLAVE $[0] \leftarrow x$ y hacer $j \leftarrow n$", el paso A2 podría suprimirse, y, en muchas máquinas, la exploración sería del orden de un 20 por ciento más rápida. Para desdicha de los programadores, los lenguajes de computador más utilizados (FORTRAN normal y COBOL) no consienten el uso de 0 como índice de un registro de memoria: no es,



EL ALGORITMO DE EXPLORACION SECUENCIAL (algoritmo A) **busca una palabra dada como "entrada" en una tabla cuyos términos no siguen orden prefijado. Esta tabla tiene 25 entradas o claves: CLAVE [1], CLAVE [2], y sucesivas, hasta CLAVE [25]. Cada clave es un apellido. Supongamos que la entrada sea "Grant". El algoritmo A explora la tabla, en busca de "Grant", comparándola primero con CLAVE [25], "Wilson", después con CLAVE [24], "Taft", etcétera. El algoritmo da la salida "17", porque "Grant" resulta ser CLAVE [17]** *(arriba).* **Si la entrada hubiese sido "Gibbs", el algoritmo A habría comparado "Gibbs" con todas las claves, y la salida habría sido 0** *(abajo).*

pues, tan sencillo mejorar el algoritmo A cuando haya de formularse en un programa escrito en esos lenguajes.

El algoritmo A resuelve el problema de exploración de una tabla de palabras; pero la solución no es muy buena, a menos que el número de palabras sea bastante pequeño, del orden de 25 o menor. Si n fuese del orden de un millón, la búsqueda secuencial resultaría, por lo común, insoportablemente lenta. No tiene sentido construir una tabla tan grande y onerosa si no va a utilizarse con frecuencia; tampoco tiene sentido derrochar tiempo en explorarla. El algoritmo A equivale a buscar un número de la guía telefónica repasándola página tras página, columna por columna, mirando solo una línea cada vez. Obviamente, sabemos hacerlo mejor.

Ventajas de ordenar la tabla

De hecho, la guía telefónica constituye un instructivo ejemplo de una gran tabla de información. Si se nos pidiera hallar el número del teléfono de una persona que residiera en el número 1642 de la calle Espronceda no tendríamos más remedio que realizar una exploración secuencial, equivalente al algoritmo A, con una guía telefónica ordinaria que no estuviese organizada para la búsqueda a partir de la dirección del abonado. Por otra parte, cuando se busca el nombre de un abonado en la guía, se aprovecha su ordenación alfabética. Esta ordenación alfabética es una gran ventaja, pues una ojeada en cualquier lugar de la guía permite eliminar definitivamente muchos nombres.

Existen muchos métodos de diseñar procedimientos de exploración eficiente de tablas ordenadas con criterio sistemático. El procedimiento más sencillo comienza por examinar en primer lugar el centro de la tabla. Si la palabra x buscada es alfabética o numéricamente menor que la central, puede eliminarse la segunda mitad de la tabla; análogamente, si x es mayor que el término central, se elimina la primera mitad. Así pues, una simple comparación permite reducir a la mitad el tamaño del problema. La misma técnica puede aplicarse a la mitad restante de la tabla, y así sucesivamente, hasta hallar la palabra x requerida. Este proceso suele denominarse exploración binaria.

Aunque las ideas subyacentes al proceso de exploración binaria son sencillas, al escribir el algoritmo se requiere poner algún cuidado. En primer lugar, en una tabla que contenga un número par de elementos, no hay uno, sino dos, términos "centrales". En segundo lugar, no es evidente cuándo hay que finalizar la búsqueda si ésta va a ser infructuosa. Los docentes de ciencias de la computación han observado, en efecto, que la primera vez que se les pide a los alumnos que preparen un proceso de exploración binaria, aproximadamente el 80 por ciento de ellos cometen errores, ¡aun cuando ya han tenido más de un año de experiencia en programación! El lector que crea haber comprendido bien los algoritmos, y que nunca haya escrito ni estudiado un algoritmo de exploración binaria podría, quizás, pasar un rato agradable tratando de construir uno, antes de leer la siguiente solución:

Algoritmo B: exploración binaria. En este algoritmo se usará la misma notación que en el algoritmo A. Además se supondrá que la primera palabra, CLAVE [1], es menor que la segunda, CLAVE [2]; ésta, menor que la tercera, CLAVE [3]; y así sucesivamente, hasta la última palabra, CLAVE [n]. Esta hipótesis puede expresarse mediante CLAVE [1] < CLAVE [2] < CLAVE [3] <...< CLAVE [n].

B1. [Partida]. Poner $l \leftarrow 0$, $r \leftarrow n+1$. (Las letras l y r representan los límites izquierdo y derecho del campo a explorar, respectivamente. Más exactamente, CLAVE [j] no puede ser igual a la palabra x dada, a menos que el índice j sea mayor que l y menor que r.)

B2. [Elección del punto medio]. Poner $j \leftarrow \lfloor (l+r)/2 \rfloor$. (En esta expresión, los corchetes significan "tómese la parte entera cociente". Por tanto, si $(l+r)$ es par, se pone en j el valor $(l+r)/2$; si $(l+r)$ es impar, j se sustituye por $(l+r-1)/2$.)

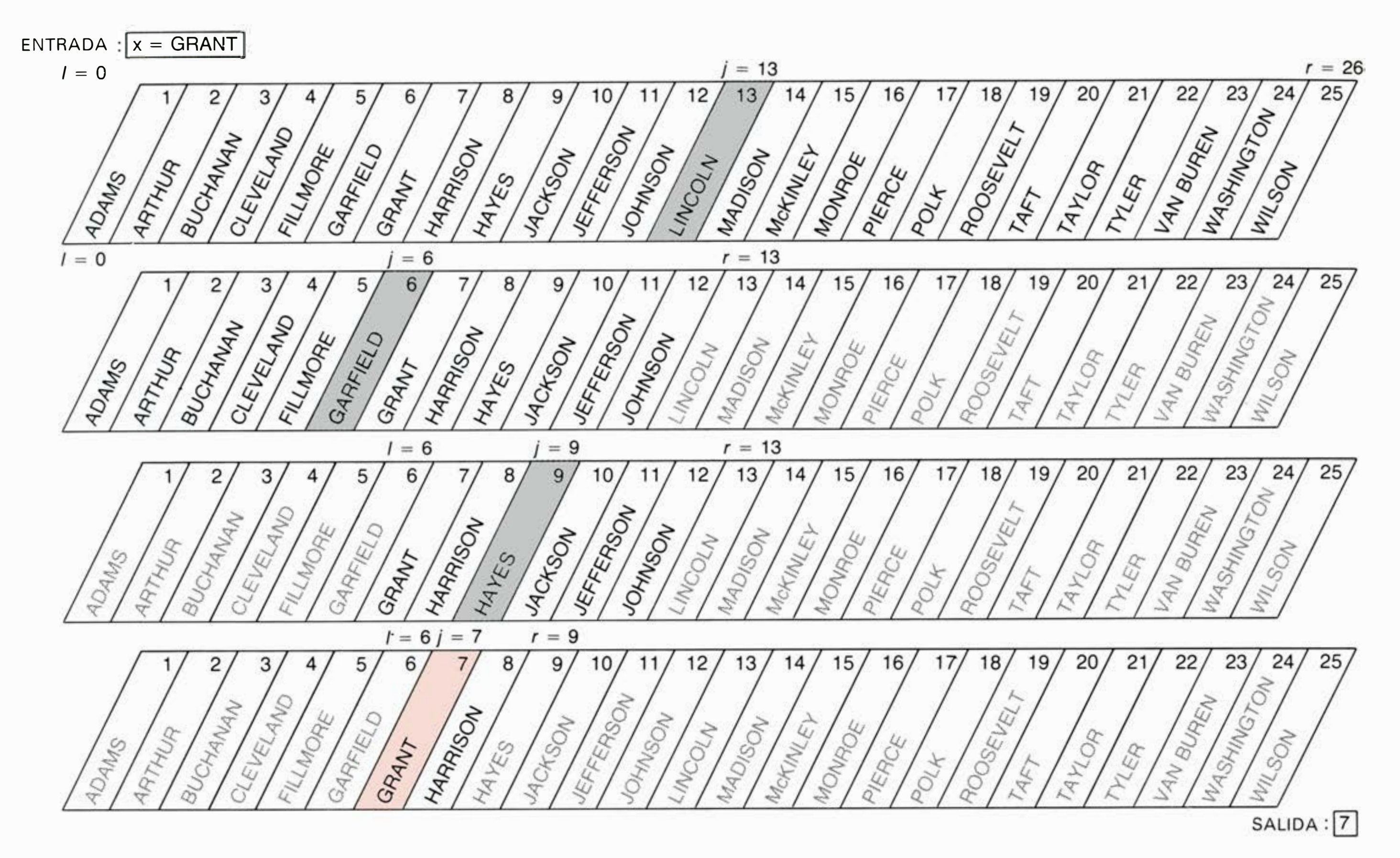
B3. [¿Fracaso?] Si $j=l$, dar salida 0 y finalizar el algoritmo. (Si j es igual a l, entonces r ha de ser igual a $l+1$ pues r siempre es mayor que l; así pues, x no puede coincidir con ninguna de las claves de la tabla.)

B4. [Comparación.] (En este punto, $j>l$ y $j<r$.) Si $x=$CLAVE [j], dar salida j y finalizar el algoritmo. Si $x<$CLAVE [j] poner $r \leftarrow j$ y regresar al paso B2. Si $x>$CLAVE [j], poner $l \leftarrow j$ y regresar al paso B2.

En la ilustración de la página siguiente se expone el proceso de exploración de

PARTIDA
A1. $j \leftarrow n$
A2. $j = 0$?
SI
NO
A3. $x=$CLAVE [j] ?
SI
SALIDA
NO
A4. $j \leftarrow j-1$
FINAL

EL DIAGRAMA DE FLUJO del algoritmo A ilustra el camino lógico por el que una exploración secuencial busca una entrada x en una tabla de n claves. El algoritmo explora la tabla comparando primero x con CLAVE [n], después, con CLAVE [n-1], luego con CLAVE [n-2], y así sucesivamente. Si x coincide con cierta CLAVE [j], el algoritmo da salida j, lugar en el que se halló x. Si x no se encuentra en la tabla, la salida del algoritmo es 0. La flecha del paso A1 ($j \leftarrow n$) significa "Hacer j igual a n" en ese paso. Los pasos de cada caja se explican detalladamente en el texto, juntamente con la descripción completa del algoritmo A. En promedio, el algoritmo A necesita examinar la mitad de la tabla para hallar x. En el caso más desfavorable, si x está en CLAVE [1], o no figura en la tabla, el algoritmo A tiene que examinarla completamente.



EL ALGORITMO DE EXPLORACION BINARIA (algoritmo B) supone un progreso importante respecto al algoritmo A, cuando la tabla a explorar es de gran tamaño. Hay que ordenar los términos de la tabla. Los 25 apellidos están ordenados alfabéticamente. La palabra "entrada" vuelve a ser "Grant". El algoritmo compara en primer lugar "Grant" con la clave que ocupa la posición central, *j*, de la tabla. Para calcular el valor de *j*, sitúa el límite izquierdo del campo a explorar *l* en 0, y el límite derecho, *r*, en $n+1$. En este ejemplo, *r* es 26. Después, la suma de *l* y *r* se divide por dos, redondeando a entero por defecto si el cociente no es entero. El punto medio de esta tabla *j*, es 26/2, o sea, 13, y está ocupado por "Lincoln" *(arriba)*. Como "Grant" es alfabéticamente menor que "Lincoln", el algoritmo desecha toda la mitad derecha de la tabla, que contiene los nombres mayores o iguales que "Lincoln". A continuación el algoritmo determina el nuevo punto medio de la mitad de tabla restante, haciendo primero *r* igual a la posición *j* recién examinada, que es 13 *(segunda figura desde arriba)*. El nuevo punto medio es $(0+13)/2$, que debe redondearse a 6, donde se encuentra "Garfield". "Grant" es alfabéticamente mayor que "Garfield", por lo que se desecha la cuarta parte izquierda de la tabla, y el límite izquierdo *l* se sitúa en 6 *(segunda figura desde abajo)*. Repitiendo una vez más el proceso, se encuentra "Grant" en la posición 7 *(abajo)*. Si la palabra "entrada" hubiese sido "Gibbs", el algoritmo habría ejecutado un paso más, con *l* todavía a 6 y *r* situado en 7. El punto medio *j* habría sido 6, límite izquierdo del campo.

una tabla de 25 nombres utilizando el algoritmo B.

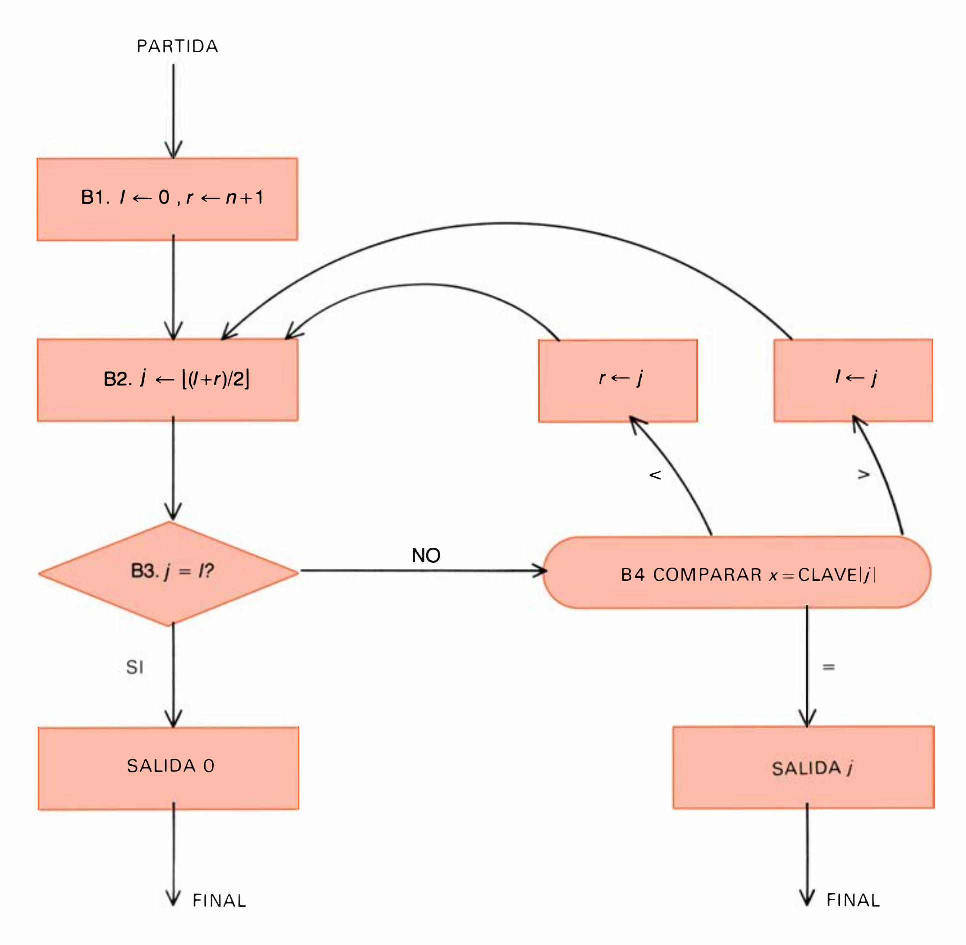
Parece obvio que la exploración binaria (algoritmo B) es muy superior a la secuencial (algoritmo A). ¿Cómo medir esta superioridad? ¿Cuándo es efectivamente superior? Tales preguntas exigen un análisis cuantitativo.

Análisis cuantitativo

Examinemos en primer lugar los casos más desfavorables para los algoritmos A y B. ¿Cuánto puede tardar cada algoritmo en hallar la palabra x en una tabla de tamaño n? Para el algoritmo A la solución es muy sencilla. Si x es igual a CLAVE [1] o si x no consta en la tabla, serán necesarias n repeticiones del paso A3; es decir, la palabra buscada habrá de compararse con la totalidad de las n entradas en la tabla antes de que la búsqueda pueda darse por finalizada. Por otra parte, el algoritmo nunca ejecutará el paso A3 más de n veces. Si se explora una tabla de un millón de registros con el algoritmo A, en el caso más desfavorable serán necesarias un millón de comparaciones.

Para la exploración binaria la solución es tan sólo un poco más difícil. Como el algoritmo B elimina la mitad de la tabla restante tras cada ejecución del paso B4, al principio ha de manejar toda la tabla; después, la mitad, la cuarta parte, la octava parte, y así sucesivamente. El número máximo de iteraciones del paso B4 será k, mínimo exponente entero tal que 2^k sea mayor que n. Por ejemplo, en la exploración binaria de una tabla de un millón (10^6) de palabras, k será igual a 20, pues 2^{20} es mayor que 10^6, mientras que 2^{19} es menor que 10^6. Así pues, en la exploración de una tabla de 10^6 palabras por medio de algoritmo B, será necesario examinar, a lo sumo, 20 entradas de la tabla, para cualquier exploración determinada.

Contemplando tan sólo el rendimiento en el caso más desfavorable, podemos ir más lejos todavía: el algoritmo B no es sólo un buen algoritmo de búsqueda; en realidad, es el óptimo de todos los algoritmos que operen por comparación de x con claves de la tabla. Ello es debido a que ningún algoritmo basado en comparaciones puede en modo alguno examinar más de $2^k - 1$ claves diferentes en sus k primeras comparaciones. Cualquiera que sea la estrategia que se adopte, para la primera comparación se ha de seleccionar determinada clave de la tabla; la segunda comparación se hará con otras dos claves a lo sumo (según que x haya sido menor o mayor que la primera palabra elegida en la tabla), la tercera comparación se hará con otras cuatro claves, a lo sumo; la cuarta, con ocho, y así sucesivamente. Por consiguiente, si en un algoritmo de búsqueda por comparación no se realizan más de k comparaciones, la tabla no puede contener más



EL DIAGRAMA DE FLUJO del algoritmo B ilustra las reglas que gobiernan la exploración binaria. El algoritmo busca la palabra *x*, "entrada", en una tabla de *n* claves que previamente han sido ordenadas. Primero se compara *x* con la clave central de la tabla. Si *x* es mayor (>) que ella, *x* se compara con el término central de la mitad derecha de la tabla; si *x* es menor (<), se compara con el centro de la mitad izquierda. El proceso continúa, desechándose cada vez la mitad de la tabla, hasta que se encuentra *x*, o la exploración prueba que *x* no está en la tabla. En nuestro simbolismo, los semicorchetes (⌊ ⌋) significan "redondear a entero por defecto".

de $1+2+4+8+\ldots+2^{k-1}$ claves distintas, y esta suma es igual a $2^k - 1$.

El conocido juego de veinte preguntas puede analizarse razonando en forma análoga. En este juego, uno de los participantes escribe en un papel el nombre de un objeto que ha pensado, plegándolo después para mantenerlo en secreto. Los restantes jugadores tratan de adivinar de qué objeto se trata, pudiendo hacer un máximo de 20 preguntas en total, a las que habrá de responderse "sí" o "no". Al principio del juego, se dice también si el objeto es animal, vegetal o mineral, o combinación de estos atributos, supuestamente bien definidos. Razonando como se ha hecho en el párrafo precedente, puede demostrarse que los restantes jugadores no pueden, de ninguna forma, identificar correctamente más de 2^{23} objetos diferentes, por muy agudas que sean sus preguntas. Existen solamente 2^3 (es decir, ocho) posibles subconjuntos del conjunto de atributos "animal, vegetal o mineral", y solamente hay 2^{20} posibles resultados de las 20 preguntas "sí" o "no". Por tanto, el número total de objetos que se pueden identificar es 2^{23}. El razonamiento es valedero aún cuando cada pregunta dependa de las respuestas a las anteriores.

Enunciemos esta conclusi6n de otra forma. Si se han de identificar más de 2^{23} objetos diferentes, no siempre serán suficientes 20 preguntas. El problema de búsqueda es parecido, pero no completamente idéntico, pues un algoritmo de exploración no se limita a preguntar "sí" o "no"; las preguntas que plantean los algoritmos de los tipos que estamos considerando tienen tres posibles respuestas, a saber: $x <$ CLAVE $[j]$, o bien $x =$ CLAVE $[j]$ o bien $x >$ CLAVE $[j]$. Si una tabla contiene 2^k o mayor número de entradas, el razonamiento precedente muestra que k comparaciones de x con claves de la tabla no serán siempre suficientes. Todo algoritmo que explore una tabla de un millón de palabras mediante comparaciones habrá de examinar, en algunos casos, 20 o más de esas palabras. En resumen, la exploración binaria es óptima cuando se da el caso pésimo.

El comportamiento de un algoritmo en los casos más desfavorables no lo es todo; tomar decisiones basándose exclusivamente en lo que puede suceder en el peor de los casos es de un pesimismo superlativo. Puede obtenerse una comprensión más adecuada de los méritos relativos de los algoritmos A y B analizando sus eficiencias en casos promedios. ¿Cuál será, por término medio, el número de comparaciones necesarias para hallar una palabra, suponiendo que las n claves de la tabla sean equiprobables? Para la exploración secuencial, la respuesta es la media aritmética simple, $(1+2+3+\ldots+n)/n$, que es igual a $(n+1)/2$. Con otras palabras, para hallar x con el algoritmo A se tendrá que explorar, por término medio, alrededor de la mitad de la tabla. Determinar el número promedio de comparaciones necesarias para hallar x mediante exploración binaria (algoritmo B) es sólo un poco más complicado. En este caso, resulta ser $k-\lfloor(2^k-k-1)/n\rfloor$, siendo k, al igual que antes, el número mínimo de comparaciones necesarias en el caso más desfavorable. Para valores grandes de n, la fórmula anterior vale, aproximadamente, $k-1$; así pues, en el caso promedio, el algoritmo B requiere tan sólo un ensayo menos que en su caso más desfavorable. Repitiendo, con las correcciones necesarias, para el caso promedio el razonamiento ya realizado para el caso más desfavorable, puede demostrarse que la exploración binaria es también el algoritmo óptimo para el caso promedio: todo algoritmo de búsqueda por comparación debe efectuar, en promedio, al menos $k-\lfloor(2^k-k-1)/n\rfloor$ comparaciones, y, en muchos de ellos, son necesarias muchas más.

Mejor que el óptimo

Tan pronto como se demuestra que algo es imposible, muchas personas intentan al punto realizarlo. Al parecer, nos hallamos en este respecto ante un factor inherente al comportamiento humano. Como acabo de demostrar que la exploración binaria es la manera óptima de explorar la memoria de un computador, nada, pues, más natural que intentar hallar un método superior.

En primer lugar, cuando el número de palabras de la tabla es pequeño, el algoritmo A resulta en realidad preferible al algoritmo B. ¿Por qué este resultado no contradice la demostración de superioridad de la exploración binaria? La razón es que al comparar el algoritmo A con el algoritmo B, hasta ahora sólo se ha

tenido en cuenta el número de comparaciones que ha de ejecutar cada uno de ellos. En realidad, el algoritmo A exige al computador menos trabajo de archivo, por lo que ejecuta cada comparación con mayor celeridad. En un computador típico, el algoritmo A viene a necesitar unas $2n+6$ unidades de tiempo, en promedio, para una tabla de tamaño n. Por otra parte, el algoritmo B requiere un promedio de, aproximadamente, $12 \log_2 n - 11 + 12(k+1)/n$ unidades de tiempo, en las mismas circunstancias. Así que, salvo cuando haya 20 o más registros que examinar, el algoritmo A será superior al algoritmo B. Estas cifras varían ligeramente de un computador a otro, pero sirven para demostrar que la eficiencia de un algoritmo no puede determinarse contando tan sólo el número de comparaciones que realiza.

Hay otra razón por la que el algoritmo B puede verse superado. Cuando al buscar un nombre en la guía telefónica se compara el nombre x buscado con los de cierta página, nuestra conducta subsiguiente no está condicionada únicamente porque tal comparación muestre que el nombre requerido sea alfabéticamente anterior o posterior a los de la página: también observamos cuán anterior o posterior es, y volvemos más o menos páginas en consecuencia. Las anteriores demostraciones de superioridad de la exploración binaria no incluyen a los algoritmos que toman en consideración nociones como el grado de diferenciación entre x y una determinada clave. La demostración concerniente a las veinte preguntas puede criticarse con similar fundamento. Por ejemplo, los jugadores podrían observar la longitud de la palabra secreta al tiempo de escribirla, o podrían obtener información a partir de la duración de los momentos de duda del jugador antes de responder "sí" o "no".

Así pues, por muy eficaz que se quiera ser, no parece necesario consultar la guía telefónica por bisecciones sucesivas, como haría un computador. Posiblemente, el método tradicional, consagrado por la experiencia, consistente en interpolar con ayuda de la ordenación alfabética, sea mejor, a pesar de la demostración de que la exploración binaria es óptima. De hecho, Andrew C. Yao, del Instituto de Tecnología de Massachusetts y F. Frances Yao, de la Brown University, han demostrado recientemente que el número de veces que, en promedio, un algoritmo de interpolación necesita acceder a la tabla es solamente $\log_2 \log_2 n$ más, a lo sumo, una constante pequeña, en el supuesto de que las entradas de la tabla sean números aleatorios independientes y uniformemente distribuidos. Cuando n es muy grande, $\log_2 \log_2 n$ es mucho menor que $\log_2 n$, por lo que la búsqueda por interpolación será significativamente más rápida que la binaria. La idea subyacente en la demostración de los Yao es que cada iteración de una búsqueda por interpolación tiende a reducir la incertidumbre en la localización de z desde n a la raíz cuadrada de n. Además, han demostrado que la búsqueda por interpolación es casi la mejor posible, en un sentido muy amplio: cualquier algoritmo que explore una tal tabla de números aleatorios efectuando las oportunas comparaciones, ha de examinar, en promedio, aproximadamente $\log_2 \log_2 n$ términos.

Estos resultados revisten gran importancia teórica; pero la experiencia en computación ha mostrado que, en la práctica, la búsqueda por interpolación no suele ser superior a la exploración binaria. El motivo es que, por lo general, los datos almacenados en una tabla real no son lo suficientemente aleatorios como para poderlos suponer uniformemente distribuidos. Además, lo corriente es que n sí sea lo bastante pequeño como para que el cálculo extra por cada comparación, exigido por cada interpolación, supere el tiempo total ahorrado al disminuir el número de comparaciones. La sencillez de la exploración binaria es una de sus virtudes, y conviene, además, guardar el debido equilibrio entre teoría y práctica.

El árbol de exploración binaria

La exploración binaria puede mejorarse de otro modo: suprimiendo la hipótesis de que la probabilidad de ser buscada es la misma para todas las claves de la tabla. Si se sabe que se requieren ciertas claves con bastante más frecuencia que otras, un algoritmo eficiente las examinará en primer lugar.

Antes de analizar esta noción, convendrá contemplar la exploración binaria bajo otro ángulo. Consideremos las 31 palabras más frecuentemente utilizadas en el idioma inglés (de acuerdo con Helen Fouché Gaines, en su libro *Cryptanalysis*). Una vez dispuestas alfabéticamente estas palabras en los registros CLAVE [1], CLAVE [2], CLAVE [3],..., CLAVE [31] de una tabla, el algoritmo B compara en primer lugar la palabra deseada x con la central, CLAVE [16], que es la palabra "I". Si x es alfabéticamente anterior a "I", la próxima comparación se realizará con CLAVE [8], que es la palabra "by"; si x es posterior a "I", la comparación siguiente se hará con CLAVE [24], que es "that". El algoritmo B opera sobre la tabla de palabras siguiendo una estructura que se asemeja a un árbol invertido, comenzando por lo más alto y descendiendo hacia la izquierda, cuando x es menor, y hacia la derecha, cuando x es mayor *(véase la ilustración superior de la página 50)*. No es difícil comprender que cualquier algoritmo ideado para explorar sólo mediante comparaciones una tabla ordenada puede describirse con auxilio de un árbol binario parecido.

El árbol de exploración binaria está implícitamente definido en el algoritmo B, mediante operaciones aritméticas con l, r y j. También puede definirse explícitamente almacenando la información concerniente al árbol en la propia tabla de palabras. A tal fin, denotemos IZQUIERDA [j] el registro de la tabla que deberá examinarse cuando x sea menor que CLAVE [j], y sea DERECHA [j] el registro a examinar si x es mayor que CLAVE [j]. Por ejemplo, la exploración binaria de una tabla de 31 palabras daría IZQUIERDA [16] igual a 8, y DERECHA [16] igual a 24, pues la búsqueda comienza en CLAVE [16] y prosigue después, bien a CLAVE [8], bien a CLAVE [24]. Si la búsqueda debe concluir infructuosamente en el paso siguiente al de comparación de la palabra x con CLAVE [j], se hace, respectivamente IZQUIERDA [j] igual a cero, 0, o DERECHA [j] igual a 0, según que x fuera mayor o menor que CLAVE [j]. En las ilustraciones de la página 50, estos ceros se representan por pequeños nódulos cuadrados, al pie del árbol.

En un árbol binario, el emplazamiento o registro de la primera de las claves por examinar se conoce tradicionalmente con el nombre de "raíz"; en el ejemplo de las 31 palabras, la raíz es 16. Pueden construirse algoritmos de búsqueda que no comiencen examinando CLAVE [16]; algoritmos que pueden ser mucho más eficientes que el algoritmo B si ciertas palabras son solicitadas con mucho mayor frecuencia que otras. He aquí una generalización de la exploración arborescente:

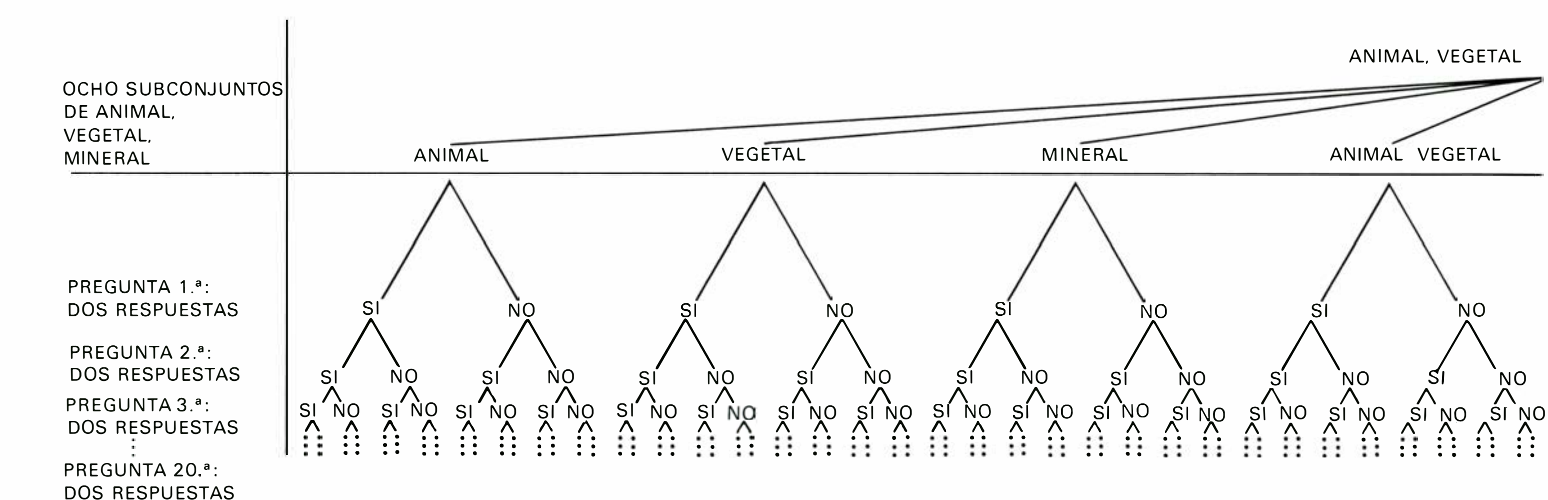
Algoritmo C; exploración arborescente.

C1. [Partida]. Hacer j igual al índice del registro de la raíz del árbol de exploración binaria.

C2. [¿Fracaso?] Si $j=0$, dar salida 0 y finalizar el algoritmo.

C3. [Comparación.] Si $x=$ CLAVE [j], dar salida j y finalizar el algoritmo. Si $x <$ CLAVE [j], poner $j \leftarrow$ IZQUIERDA [j] y regresar al paso C2. Si $x >$ CLAVE [j], poner $j \leftarrow$ DERECHA [j], y regresar al paso C2.

El algoritmo C se asemeja a un texto



EL JUEGO DE LAS VEINTE PREGUNTAS pone de relieve una fundamental limitación de cualquier método de exploración arborescente. Un jugador piensa un objeto, del que dice si es animal, vegetal o mineral, o cualquier combinación de estas características. Sus oponentes tratan de adivinar de qué objeto se trata, haciendo un máximo de 20 preguntas, que han de responderse con un "sí" o con un "no". Los jugadores no lograrán identificar más de 2^{23}, o sea, 8.388.608, objetos diferentes. La razón estriba en que el conjunto de características animal, vegetal y mi-

de enseñanza programada, en el que según se responda a ciertas preguntas, cada página va diciéndole al lector a qué nueva página debe dirigirse. El algoritmo funciona en cualquier árbol binario en el que, para todos los valores de *j* del árbol, todas las claves accesibles desde IZQUIERDA [*j*] sean menores que CLAVE [*j*], y todas las claves accesibles desde DERECHA [*j*] sean mayores que CLAVE [*j*]. Un árbol así se llama árbol de exploración binaria.

Una de las ventajas de algoritmo C sobre el algoritmo B es que no requiere cálculo aritmético alguno, por lo que el computador realiza la búsqueda algo más rápidamente. Sin embargo, la ventaja principal del algoritmo C es que la estructura arborescente consiente mayor flexibilidad, pues las palabras de la tabla pueden ahora redisponerse en un orden cualquiera. Ya no es necesario que CLAVE [1] sea menor que CLAVE [2], y así sucesivamente hasta CLAVE [*n*]. Mientras los marcadores IZQUIERDA y DERECHA definan un árbol de exploración binaria válido, las posiciones reales de las claves de la tabla carecen de importancia.

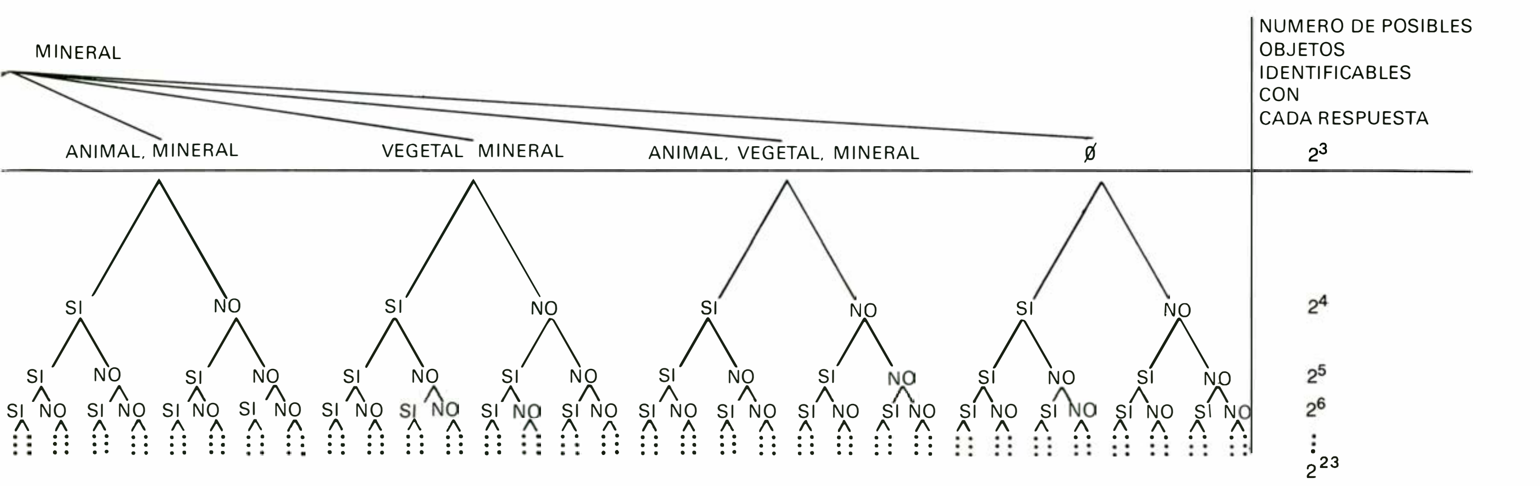
Esto significa, pues, que pueden añadirse nuevas palabras, sin necesidad de alterar la posición de las anteriores. Por ejemplo, podría añadirse la palabra "has" al árbol binario de las 31 palabras sin más que poner CLAVE [32]←"has" y cambiar DERECHA [*j*] de 0 a 32, siendo *j* el emplazamiento de la clave "has". Podría pensarse que tales añadiduras al pie del árbol perturbarán gravemente el equilibrio de la estructura, pero puede demostrarse matemáticamente que si se añaden al azar nuevas palabras, el resultado será, con casi total seguridad, un árbol razonablemente equilibrado.

Arboles óptimos de exploración binaria

Como el algoritmo C es válido para cualquier árbol de exploración binaria, podemos "cortar" el árbol a la medida, con el fin de que las claves más solicitadas sean examinadas en primer lugar. Se consigue así abreviar el tiempo medio que necesita el computador para realizar la búsqueda, aunque no el tiempo del caso más desfavorable. La ilustración inferior de la página 50 muestra el árbol binario óptimo para las 31 palabras inglesas más corrientes, basándose en la estimación de frecuencias realizada por Gaines. En promedio, el número de comparaciones necesarias para hallar *x* en este árbol óptimo binario es de tan sólo 3,437, mientras que en el árbol equilibrado este número asciende a 4,393. Vale la pena notar que el árbol óptimo, basado en la frecuencia de las palabras, no comienza comparando *x* con la palabra "the". A pesar de que la palabra "the" es con mucho la más corriente del idioma inglés, su retrasada posición en el orden alfabético la aparta demasiado del centro para ser raíz óptima.

Desde el punto de vista de la matemática convencional, el problema de hallar el árbol binario óptimo es trivial, pues para cada conjunto particular de *n* palabras y sus correspondientes frecuencias existe tan sólo un número finito de árboles de exploración binaria. En principio, bastaría catalogar todos los posibles y elegir el que mejor funcione; en la práctica, sin embargo, tal observación es perfectamente inútil, porque el número de árboles binarios de *n* elementos es $(2n)!/n!\ (n+1)!$, donde $n!$ representa el producto $1\times2\times3\times...xn$. La fórmula muestra que, de hecho, hay muchísimos árboles binarios, unos $4^n/\sqrt{(\pi\ n^3)}$, siendo π el conocido 3,14159. Así, por ejemplo, cuando *n* es 31, el número total de posibles árboles binarios es 14.544.636.039.226.909, y cada uno de estos catorce mil billones de árboles será el árbol óptimo para cierto conjunto de presuntas frecuencias de las 31 palabras. ¿Cómo es entonces posible demostrar que el árbol presentado como óptimo para las frecuencias de Gaines lo es verdaderamente? El más rápido de los modernos computadores está muy lejos de poder examinar los catorce mil billones de casos particulares: si se estudiase un árbol cada microsegundo, la tarea ocuparía 460 años.

Existe, sin embargo, un importante principio que hace realizable el cómputo: todo subárbol de un árbol óptimo también ha de ser óptimo. En el árbol óptimo para las 31 palabras inglesas más comunes, el subárbol situado a la izquierda de la palabra "of" ha de proporcionar el mejor método posible de explorar las 20 palabras "a", "and", y sucesivas, hasta "not". Pues si hubiera un método superior, insertándolo en el árbol total se tendría un árbol total mejor, con lo que el árbol dado no sería el óptimo. Análogamente, en dicho subárbol, el árbol menor situado a la derecha de "for" ha de ser el mejor modo posible de explorar las 11 palabras "from", "had" y sucesivas, hasta "not". Cada subárbol se corresponde con un conjunto de palabras consecutivas CLAVE [*i*], CLAVE [*i*+1], ..., CLAVE [*j*], en donde $1\leq i\leq j\leq n$. Pue-



neral tiene solo ocho, o sea, 2^3, posibles subconjuntos (contando también el conjunto vacío ∅); estas ocho posibilidades se combinan con 2^{20} posibles resultados de las 20 preguntas de respuesta "sí" o "no". Un razonamiento parecido puede servir para demostrar que un algoritmo de búsqueda que formule a lo sumo 20 preguntas del tipo "menor, igual, o mayor" no puede distinguir más del $2^{20}-1$ valores distintos de claves de la tabla, pues $1+2+4+8+...+2^{19}=2^{20}-1$. La exploración binaria alcanza este límite, por lo que es, aquí, el algoritmo más eficiente.

den determinarse todos los subárboles óptimos, hallando en primer lugar los más pequeños y realizando el cálculo para valores crecientes de $j-i$. Para cada elección de i y j hay $j-i+1$ posibles raíces del subárbol. Conforme avanza el cálculo y se asciende por el árbol, al examinar la raíz de cada posible subárbol, los subárboles óptimos de su derecha y de su izquierda habrán sido calculados ya.

Este procedimiento permite hallar de modo efectivo el árbol óptimo de exploración binaria para n claves y frecuencias, realizando para ello alrededor de n^3 operaciones. De hecho, el autor ha podido mejorar todavía más este método, con lo que el número de operaciones necesarias puede reducirse a n^2. En el caso de las 31 palabras, ello significa que puede determinarse el árbol óptimo después de apenas 961 etapas, en lugar de 14.000 billones.

Debe señalarse que en los párrafos precedentes se han examinado varios algoritmos cuyo único objetivo es determinar el árbol óptimo de exploración binaria. En otras palabras, ¡la salida de estos algoritmos es en sí un algoritmo, destinado a resolver otro problema! Este ejemplo contribuye a explicar por qué las ciencias de la computación se han desarrollado tan rápidamente como disciplina independiente. Al estudiar la más adecuada utilización de los computadores aparecen problemas interesantes por sí mismos, y muchos de estos problemas exigen elaborar todo un conjunto de técnicas y nociones nuevas e interconectadas.

Resulta a la vez entretenido e instructivo considerar el árbol de exploración binaria peor posible para las 31 palabras inglesas de uso más corriente, a fin de comprobar hasta qué punto pueden ir mal las cosas utilizando el algoritmo C. Al igual que con los árboles óptimos, existe un método para determinar árboles pésimos que requiere unas n^2 operaciones. En las 31 palabras, con las frecuencias dadas por Gaines, el árbol pésimo de exploración binaria exige del algoritmo C un promedio de 19,158 comparaciones por exploración. A título de comparación, la peor ordenación posible de las claves en una exploración secuencial obliga al algoritmo A a ejecutar un promedio de 22,907 comparaciones por cada exploración. Así que el peor caso del algoritmo C no será nunca tan malo como el peor caso del algoritmo A.

Condensación

Los algoritmos de exploración precedentes están íntimamente relacionados con los métodos humanos de consultar diccionarios. En realidad, existe un método muy superior de explorar, mediante computador, una gran tabla de palabras. Se llama método de condensación, y aborda el problema de modo radicalmente diferente, e imposible de usar por humanos, pues se basa en la capacidad de la máquina para efectuar cálculos aritméticos a gran velocidad. En esencia, consiste en manejar las letras de cada palabra como si fueran números ($a=1$, $b=2$, $c=3$,... hasta $z=26$), que luego se condensan en un solo número, unívocamente asociado a cada palabra. Este número se llamará "dirección condensada" de la palabra; indica al computador en qué lugar de la tabla debe buscarse la palabra.

En el ejemplo de las 31 palabras inglesas más comunes, se puede condensar cada palabra en un número comprendido entre 1 y 32 sumando los valores numéricos de sus letras y luego restando, si es necesario, los múltiplos de 32 sobrantes. Por ejemplo, la dirección así condensada de la palabra "the" sería $20+8+5-32=$ $=1$; la dirección condensada de la palabra "of" sería $15+6=21$, y así sucesivamente para el resto de la tabla. Si se tiene suerte, cada palabra tendrá una dirección condensada diferente, y las exploraciones de la tabla serán sumamente rápidas.

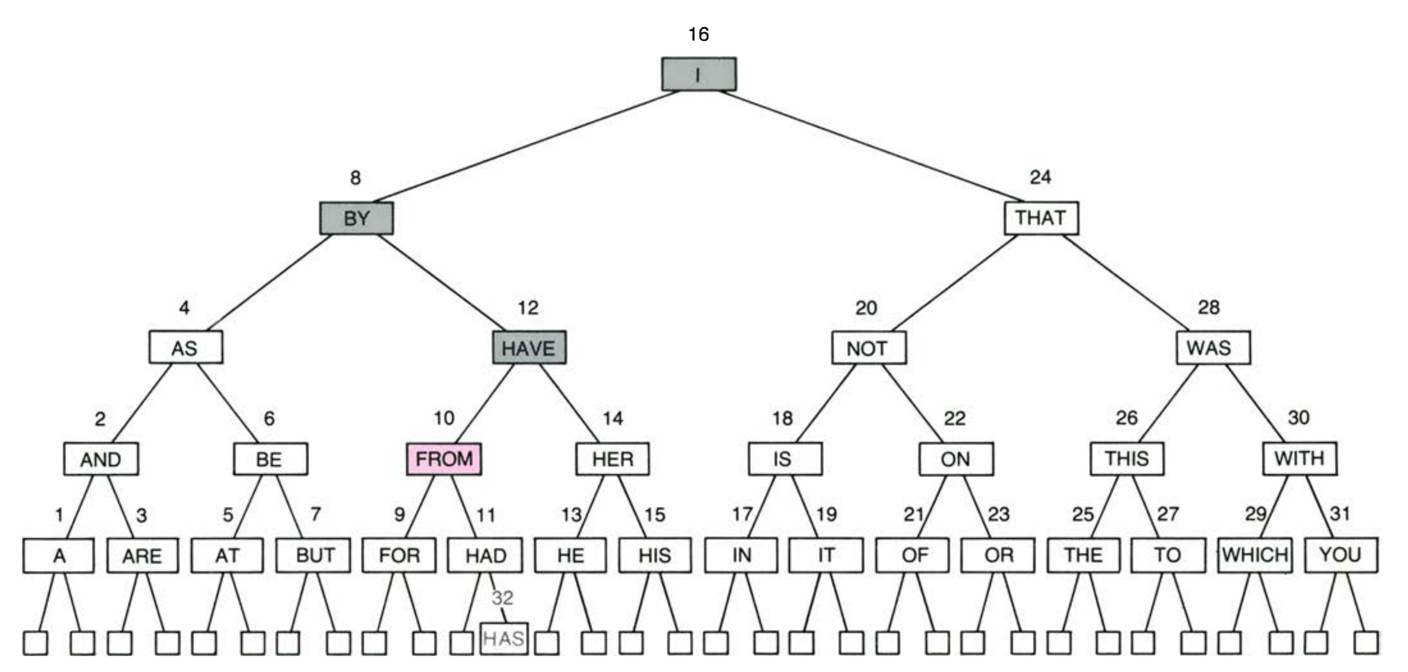
De modo general, supongamos que haya m registros en la memoria de un computador, y supongamos que se deseen almacenar n claves en ella, siendo m mayor que n. Como n es igual a 31, digamos que m sea 32. Supongamos también que haya una función de condensación $h(x)$ que convierta toda posible palabra x en un número comprendido entre 1 y m. Una buena función de condensación habrá de tener la propiedad de que raramente sean iguales $h(x)$ y $h(y)$, siendo x e y palabras distintas por insertar en la tabla.

Sin embargo, excepto en los casos en los que m sea mucho mayor que n, prácticamente todas las funciones de condensación provocarán cuando menos varias "colisiones" de valores $h(x)$ y $h(y)$. Es extremadamente improbable que n números aleatorios independientes comprendidos entre 1 y m sean todos distintos. Tomemos un ejemplo corriente: es sabido que si se reúnen en la misma sala 23, o mayor número, de personas,

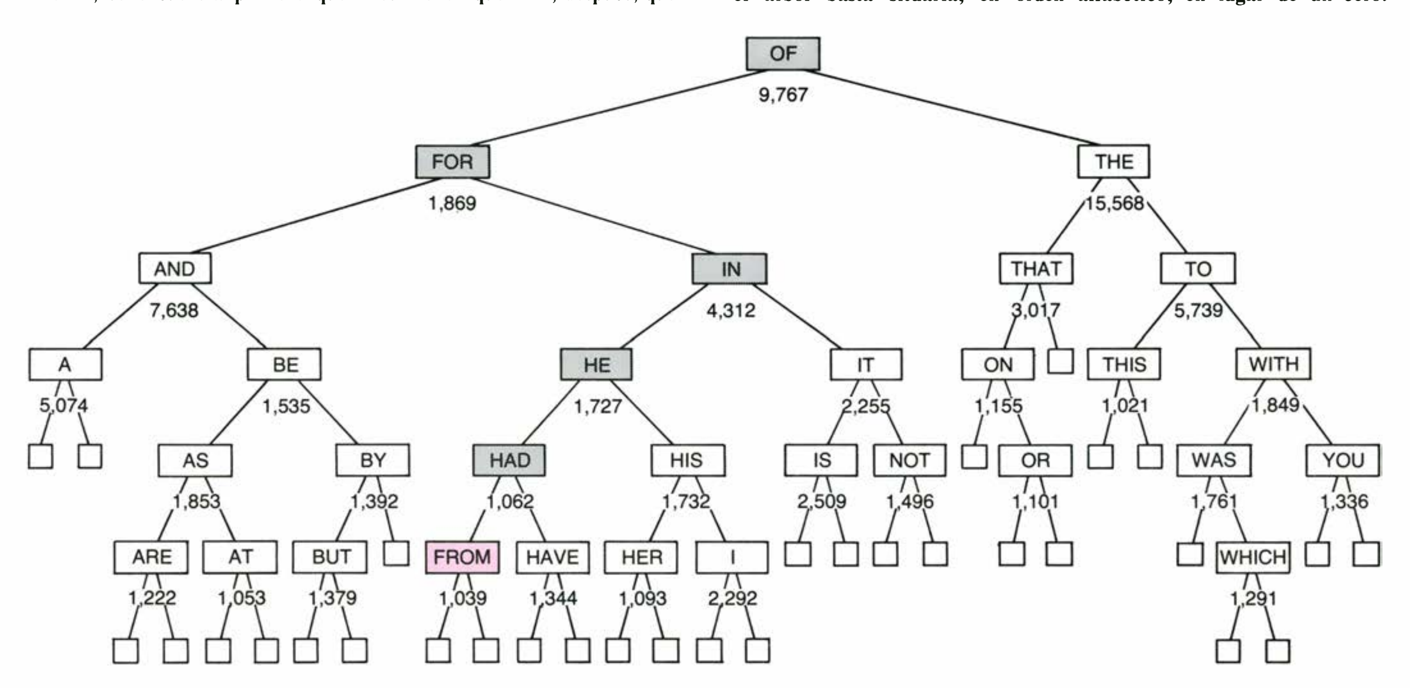
la probabilidad de que dos de ellas tengan el mismo cumpleaños es mayor que la par. Además, en un grupo de 88 personas, es probable que haya tres con el mismo cumpleaños. Aunque este fenómeno le resulta paradójico a muchos, es fácil de comprobar matemáticamente, y muchas coincidencias, que parecían imposibles, tienen explicación similar.

Otra forma de enunciar la paradoja de los cumpleaños es afirmar que, en conjunto, más de la mitad de las funciones de condensación con m igual a 365 y n igual a 23 tendrán al menos una colisión. Por consiguiente, todo proceso de exploración basado en funciones de condensación deberá poder afrontar el problema de las colisiones.

Supongamos que sea necesario buscar x en una tabla, y que su dirección condensada $h(x)$ contenga ya la palabra y.



ARBOL DE EXPLORACION BINARIA implícito en el algoritmo B. El árbol muestra gráficamente cómo iría sondeando el algoritmo B la tabla alfabéticamente ordenada de las 31 palabras inglesas más utilizadas. Comenzando en la "raíz", o vértice superior, del árbol, la palabra "entrada" x se compara primero con el centro de la tabla, la palabra "I". Si x es alfabéticamente menor que "I", la búsqueda prosigue bajando por la rama izquierda del árbol; si x es mayor, se desciende por la rama derecha. Por ejemplo, si x es la palabra "from", se encuentra primero que x es menor que "I", después, que x es mayor que "by", a continuación, que x es menor que "have", y, finalmente, que x es igual a "from". Si x no estuviera en la tabla, la búsqueda finalizaría en uno de los 32 ceros *(nódulos cuadrados)* al pie del árbol. Representando explícitamente las ramas del árbol en la memoria del computador, en lugar de hacerlo implícitamente mediante el algoritmo B (que exige calcular los puntos medios), la exploración es un poco más rápida. Resulta también más sencillo insertar nueva información: para insertar "has" *(escrita en trazos grises)* en el árbol basta situarla, en orden alfabético, en lugar de un cero.



ARBOL OPTIMO de exploración binaria que muestra la mejor ordenación posible de las 31 palabras del árbol, de acuerdo con la frecuencia relativa de cada palabra, en estimación de Helen Fouché Gaines. La frecuencia de cada palabra es el número que figura bajo ella. Este árbol no es tan armonioso como el implícitamente definido por el algoritmo binario normal, mostrado en la ilustración superior; así pues, en algunos casos, la búsqueda puede prolongarse más tiempo. Por ejemplo, para hallar la palabra "from" en este árbol son necesarios seis pasos, en lugar de cuatro *(vía de gris y de color)*. Sin embargo, en promedio, el árbol óptimo se ve explorado con mayor rapidez por el computador. A pesar de ser "the" la palabra más utilizada del inglés, no está en la raíz, porque alfabéticamente dista mucho del centro.

El modo más sencillo de tratar esta colisión consiste en examinar los registros $h(x)$, $h(x)-1$, $h(x)-2$, etc., hasta que se halle x o bien se llegue a un registro vacío.

Si en la búsqueda se sale por un extremo de la tabla antes de que la exploración haya finalizado, se recomienza por el otro extremo. Este proceso, que llamaremos "de sondeo lineal", puede formularse mediante un algoritmo:

Algoritmo D; condensación con sondeo lineal.

D1. [Partida.] Poner $j \leftarrow h(x)$.

D2. [¿Fracaso?] Si el registro j de la tabla está vacío, dar salida 0 y finalizar el algoritmo.

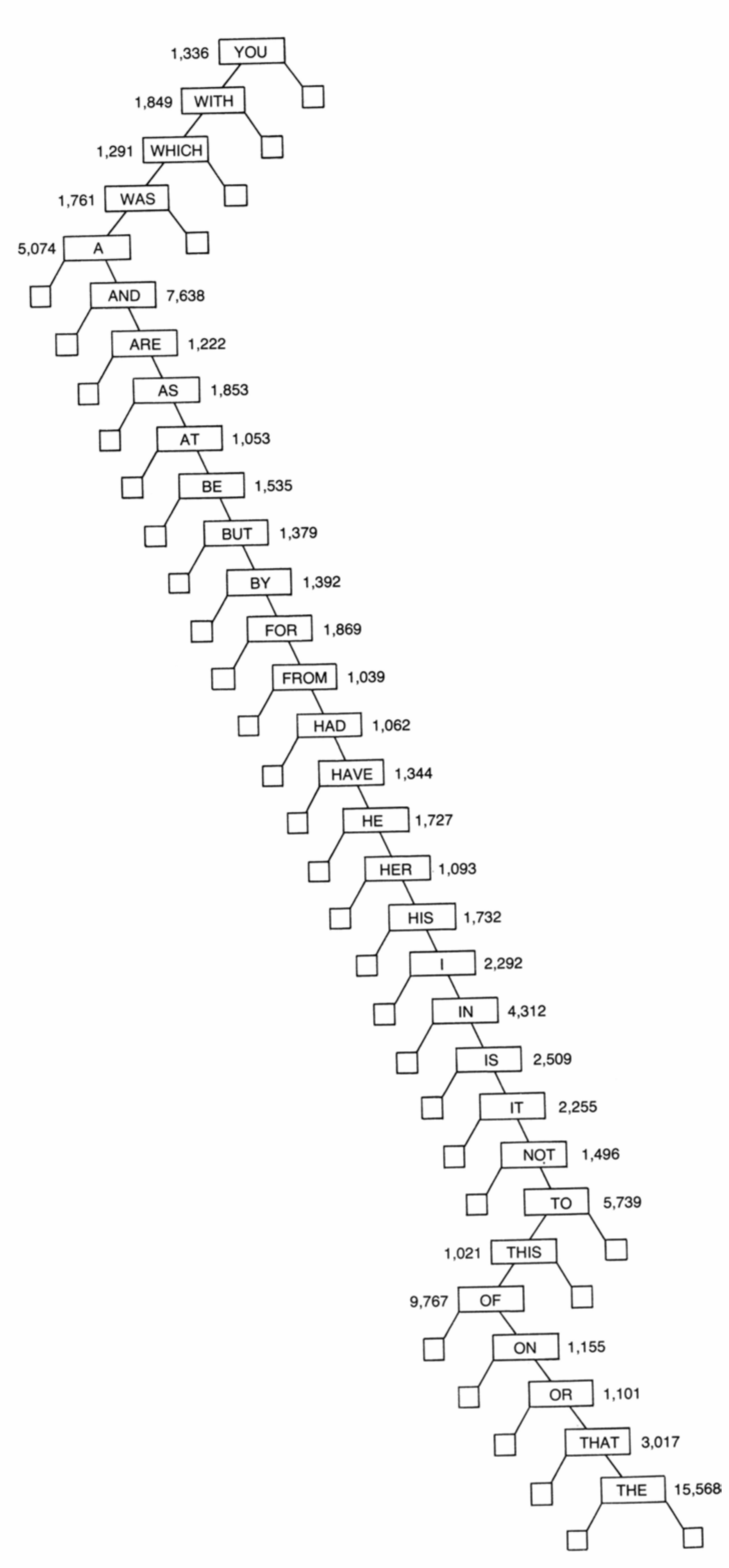
D3. [¿Exito?] Si $x=$ CLAVE $[j]$, dar salida j y finalizar el algoritmo.

D4. [Pasar al registro adyacente.] Poner $j \leftarrow j-1$; entonces, si $j=0$, poner $j \leftarrow m$. (El registro de índice m se considera adyacente al de índice 1.) Regresar al paso D2.

Si x no se encuentra en la tabla, y si el algoritmo termina infructuosamente en el paso D2 debido a que el registro j está vacío, podríamos poner CLAVE $[j] \leftarrow x$, usando este valor de j. De esta forma se insertaría x en la tabla, con lo que más tarde el algoritmo podría recuperarla. Una posterior búsqueda de x seguiría los mismos pasos que la vez primera, comenzando en el registro $h(x)$, pasando después al $h(x)-1$, y así sucesivamente, hasta hallar x en el registro j. Así pues, la exploración se desarrollará sin tropiezos, a pesar de las colisiones.

Volviendo al ejemplo de las 31 palabras inglesas más comunes, supongamos que las palabras van siendo alojadas, una por una, en una tabla inicialmente vacía, de acuerdo con lo expuesto en el párrafo anterior, y siguiendo el orden de frecuencias decrecientes. Se inserta primero la palabra "the", después "of", y así sucesivamente. Resulta la tabla de condensación que se muestra en la ilustración de la página 52. La mayoría de los registros están ocupados por palabras cuya dirección condensada coincide con el índice del registro, o es cercana a él, a excepción de las palabras insertas en último lugar.

Así, la palabra menos frecuente, "this", ha sido insertada en el registro 8, a pesar de ser 24 su dirección condensa-



ARBOL "PESIMO" que muestra el peor de los posibles árboles de exploración binaria de las 31 palabras inglesas más habituales. Este árbol ha perdido la ventaja de la estructura ramificada: una rama está siempre "muerta".

1	THE	(1)
2	HAVE	(4)
3	TO	(3)
4	HIS	(4)
5		
6	BE	(7)
7	FOR	(7)
8	THIS	(24)
9	I	(9)
10	BUT	(11)
11	WAS	(11)
12	HAD	(13)
13	HE	(13)
14	FROM	(20)
15	AT	(21)
16	NOT	(17)
17	THAT	(17)
18	WHICH	(19)
19	AND	(19)
20	AS	(20)
21	OF	(21)
22	ON	(29)
23	IN	(23)
24	ARE	(24)
25	YOU	(29)
26	BY	(27)
27	WITH	(28)
28	IS	(28)
29	IT	(29)
30	HER	(31)
31	OR	(1)
32	A	(1)

UNA TABLA DE "CONDENSACION" facilita la exploración de series de datos al computador. Con la velocidad de cálculo de la máquina, se asocia a cada palabra *x* una dirección condensada (hash address), donde ha de comenzar la búsqueda de *x*. La dirección condensada de cada una de las 31 palabras inglesas más comunes se muestra entre paréntesis, a su lado. En este ejemplo, se obtuvieron las direcciones condensadas sumando los valores numéricos de las letras ($a = 1$, $b = 2$, y así sucesivamente, hasta $z = 26$), y restando múltiplos de 32. A veces, dos palabras diferentes tienen la misma dirección condensada, por lo que "colisionan". Si *x* no está alojada en el registro $h(x)$, la búsqueda continúa hacia arriba, examinando los registros $h(x)$-1, $h(x)$-2, etc. Por ejemplo, la dirección condensada de "his" es $h+i+s$, o sea, $8+9+19-32=4$. También la dirección condensada de "have" es 4. Para buscar "have", el algoritmo examina primero el registro número 4 *(gris claro)*, después, el número 3 *(gris oscuro)* y finalmente, en el 2 *(en color)*, donde está alojada "have". Si la palabra *x* no está en la tabla, la exploración terminaría, por supuesto, en el registro vacío número 5.

da, porque los registros números 9 a 24 estaban ya ocupados cuando se buscó dónde alojarla. A pesar de estas anormalidades, el promedio de veces que el algoritmo D necesita sondear la tabla para encontrar una palabra resulta ser solamente 1,666 (menos de la mitad de comparaciones que exige en promedio el árbol óptimo de exploración binaria). Naturalmente que el tiempo necesario para computar $h(x)$ en el paso D1 debe agregarse al tiempo ocupado en sondear la tabla. Para grandes tablas de datos, el método de condensación demostrará ser significativamente más eficiente que cualquier algoritmo de comparación binaria.

En la práctica, casi nunca se rellena tanto la tabla como se ha hecho en el ejemplo. El número m de registros disponibles se elige, por lo general, lo bastante grande como para que la tabla nunca se rellene en más del 80 o 90 por ciento. Puede demostrarse que el promedio de sondeos necesarios para hallar una palabra entre otras n igualmente probables, insertas al azar en una tabla de tamaño m, es $1+[(n-1)/m+(n-1)(n-2)/m^2+(n-1)(n-2)(n-3)/m^3+\ldots]/2$. Representamos el cociente n/m mediante el símbolo α, que es la razón de llenado o "coeficiente de carga" de la tabla. Puede demostrarse que cuando n tiende hacia infinito, el número promedio de sondeos necesarios para encontrar una palabra x de la tabla tiende hacia $1+(\alpha+\alpha^2+\alpha^3+\ldots)/2$ que es igual a $[1+1/(1-\alpha)]/2$. Además, el verdadero número medio de sondeos es siempre menor que este valor límite. Así pues, cuando la tabla por explorar está rellena en un 80 por ciento, el algoritmo D ejecuta, en promedio, menos de tres sondeos por cada búsqueda con éxito.

Es importante observar que el número promedio de sondeos necesarios en una exploración con éxito, dado antes, es válido para todas las tablas que estén igualmente rellenas, y no depende del tamaño de la tabla. No puede decirse lo mismo de los algoritmos de comparación binaria, porque su tiempo promedio de funcionamiento por cada búsqueda con éxito crecerá indefinidamente conforme aumente el número n de palabras por explorar.

Mejora de búsquedas infructuosas

Las anteriores aseveraciones sobre el pequeño número de sondeos necesarios con el algoritmo D son válidas tan sólo cuando la palabra x se encuentra de hecho en la tabla. Si x no está presente en la tabla, el número de comprobaciones necesarias para cerciorarse de ello es mayor, a saber:

$$1+[2n/m+3n(n-1)/m^2+4n(n-1)(n-2)/m^3+\ldots]/2;$$

cuando n es grande, este número se aproxima a $[1+1/(1-\alpha)^2]/2$. Dicho de otro modo, por término medio, cada exploración infructuosa de una tabla grande rellena hasta un 80 por ciento, exige casi 13 comprobaciones. Además, en el ejemplo de las 31 palabras en 32 espacios, ha de observarse que todas las exploraciones infructuosas han de finalizar en el registro número 5, único vacío, independientemente de la dirección inicial $h(x)$. Situación enteramente análoga a la que se tenía con el algoritmo A, donde toda exploración infructuosa ha de terminar en la posición 0.

En 1973, O. Amble, de la Universidad de Oslo, observó que el problema de las búsquedas infructuosas podía paliarse combinando condensación y ordenación alfabética. Supongamos que las 31 palabras inglesas más habituales se inserten en la tabla en orden alfabético decreciente, y no en orden de frecuencia decreciente. Como la tabla se sondea comenzando en la dirección $h(x)$, y pasando después a $h(x)-1$, etc., todas las palabras comprendidas entre la dirección $h(x)$ y el verdadero registro ocupado por x deben ser alfabéticamente mayores que x para que no haya colisión. Así pues, la búsqueda de x puede declararse infructuosa tan pronto como se encuentre la palabra alfabéticamente menor que x. Es decir, puede usarse el siguiente algoritmo:

Algoritmo E: sondeo lineal en tabla ordenada. Este algoritmo supone que CLAVE $[j]$ es 0 cuando el registro j está vacío, y que todas las palabras x tienen un valor numérico mayor que 0.

E1. [Partida.] Poner $j \leftarrow h(x)$.

E2. [¿Fracaso?] Si CLAVE $[j] < x$, dar salida 0 y finalizar el algoritmo.

E3. [¿Exito?] Si CLAVE $[j] = x$, dar salida j y finalizar el algoritmo.

E4. [Pasar al siguiente] Poner $j \leftarrow j-1$; entonces, si $j=0$, poner $j \leftarrow m$. Regresar al paso E2.

La ventaja del algoritmo E se muestra en la tabla condensada y ordenada de la página 53. Supongamos que se necesita determinar si "has" es una de las palabras más frecuentes en inglés. Su dirección condensada es $8+1+19=28$. Con el algoritmo E, la búsqueda finaliza en seis pasos, cuando se alcanza $j=22$ ("by"), en lugar de continuar examinando la tabla hasta llegar al registro vacío correspondiente a $j=5$.

En una tabla ordenada, el número pro-

medio de sondeos por búsqueda infructuosa se reduce a

$$1 + [n/m + n(n-1)/m^2 + n(n-1)(n-2)/m^3 + \ldots]/2.$$

y este número siempre es menor que

$$[1 + 1/(1-\alpha)]/2.$$

Vemos así que el límite para una búsqueda con éxito y el límite para una exploración infructuosa son idénticos. En promedio, en una tabla ordenada, rellena al 80 por ciento de su capacidad, el algoritmo E tendrá que efectuar menos de tres sondeos, independientemente del tamaño de n.

Cuanto antecede está muy bien... si la tabla ordenada se ha constituido en orden alfabético decreciente, como se ha explicado. En la práctica, sin embargo, no siempre puede suponerse así. Por lo común, las tablas crecen dinámicamente al emplearlas, y en ellas entran al azar nuevas palabras. Aunque las estructuras de árbol binario (algoritmo C) o de método de condensación no ordenada (algoritmo D) pueden afrontar fácilmente el problema del crecimiento dinámico, la estructura de tabla ordenada (algoritmo E) no es tan obviamente adaptable. Por fortuna, hay un algoritmo muy sencillo que resuelve el problema de la inserción de nuevas palabras en una tabla ordenada con direccionamiento condensado:

Algoritmo F: inserción en tabla condensada y ordenada. Este algoritmo coloca una nueva palabra x en una tabla condensada y ordenada, redisponiendo adecuadamente los restantes términos de modo que el algoritmo E siga siendo válido.

F1. [Partida.] Poner $j \leftarrow h(x)$.

F2. [Comparación.] Si CLAVE $[j] < x$, intercambiar los valores de x y de CLAVE $[j]$. (Es decir, dar a x el anterior valor de CLAVE $[j]$ y a CLAVE $[j]$ el anterior valor de x.)

F3. [¿Efectuado?] Si $x = 0$, finalizar el algoritmo.

F4. [Pasar al siguiente.] Poner $j \leftarrow j-1$; entonces, si $j = 0$, poner otra vez $j \leftarrow m$. Regresar al paso F2.

Si se quisiera insertar la palabra "has", mediante el algoritmo F, en la tabla condensada y ordenada de las 31 palabras inglesas más habituales, el procedimiento consistiría en hacerle sitio en la posición 22, pasando para ello "by" de la posición 22 a la 18; "at", de la 18 a la posición 15; "as" pasaría de la posición 15 a la 14; "are", de la 14 a la 8, y finalmente, desplazando "and" de la posición 8 a la posición vacía número 5. En apariencia, el proceso requiere mucho trabajo, pero en realidad, sólo se tarda un poco más que en insertar "has" en una tabla desordenada usando el algoritmo D. Además, el número de palabras que deben intercambiarse, aplicando reiteradamente el paso F2, para acomodar una nueva palabra es, por término medio, $(n-1)/2m + 2(n-1)(n-2)/3m^2 + 3(n-1)(n-2)/4m^3 + \ldots$, lo que es aproximadamente igual a $1/(1-\alpha) + [\log_e(1-\alpha)]/\alpha$, siendo e la base de los logaritmos neperianos 2,71828... Así pues, la inserción de nuevas palabras mediante el algoritmo F es una tarea perfectamente razonable.

En este ejemplo específico, no debimos insertar "has" en la tabla, porque las tablas de direccionamiento condensado ("tablas condensadas") han de tener, cuando menos, una posición vacía. Por pura coincidencia, la mínima palabra en orden alfabético, "a", se encuentra en esta tabla completamente llena, por lo que el sondeo lineal con el algoritmo E seguirá funcionando correctamente en todos los casos. Sin embargo, si "a" no estuviera en la tabla, sería necesaria una posición vacía, para evitar que la búsqueda se prolongara indefinidamente en el caso de que la palabra de entrada x fuese igual a "a".

Una de las más sorprendentes propiedades de las tablas ordenadas con direccionamiento condensado es que cada una de ellas es única. Usando el algoritmo F para construir una tabla ordenada y condensada a partir de un conjunto cualquiera de palabras, se obtendrá la misma tabla independientemente del orden de inserción de las palabras. Quizá el lector encuentre entretenido demostrar por sí mismo este hecho.

Conclusión

La exposición anterior sobre los distintos modos de buscar información almacenada en la memoria de un computador tiene por objeto iluminar varias cuestiones referentes a los algoritmos en general. Como hemos visto, un algoritmo tiene que ser formulado con precisión, lo que no es tan sencillo como puede pensarse a primera vista. Cuando se intenta resolver un problema con auxilio de un computador, lo usual es que el primer algoritmo que se nos ocurra pueda mejorarse muchísimo. Las estructuras de datos, como los árboles binarios, constituyen importantes herramientas para construir algoritmos eficientes. En cuanto se comienza a investigar la rapidez de un algoritmo, o cuando se intenta hallar el algoritmo óptimo para una aplicación específica, se plantean interesantes cuestiones, y con frecuencia se descubre que sus respuestas son harto sutiles. Incluso el algoritmo "mejor posible" puede mejorarse en ocasiones si se cambian las reglas de base. Como los computadores "piensan" de distinta forma que las personas, los métodos adecuados para la mente humana no son, necesariamente, los más eficientes en una máquina.

1	THE	(1)
2	HAVE	(4)
3	TO	(3)
4	HIS	(4)
5		
6	BE	(7)
7	FOR	(7)
8	AND	(19)
9	I	(9)
10	BUT	(11)
11	WAS	(11)
12	HAD	(13)
13	HE	(13)
14	ARE	(24)
15	AS	(20)
16	NOT	(17)
17	THAT	(17)
18	AT	(21)
19	WHICH	(19)
20	FROM	(20)
21	OF	(21)
22	BY	(27)
23	IN	(23)
24	THIS	(24)
25	IS	(28)
26	IT	(29)
27	ON	(29)
28	WITH	(28)
29	YOU	(29)
30	A	(1)
31	HER	(31)
32	OR	(1)

UNA TABLA ORDENADA y "de condensación" combina la noción de condensación con las ventajas de la ordenación alfabética, y revela más rápidamente si la palabra "entrada" no consta en la tabla. Ahora, todas las palabras situadas entre la posición *h(x)* y la verdadera posición de *x* son alfabéticamente mayores que *x*. Así que las búsquedas infructuosas no han de terminar necesariamente en la posición vacía número 5, sino que terminarán tan pronto se encuentre una palabra alfabéticamente menor que *x*. Si la palabra requerida fuese "has", cuya dirección condensada es 28 *(gris claro)*, la búsqueda finalizaría al llegar a "by", en la posición 22 *(gris oscuro)*.

Biocristales

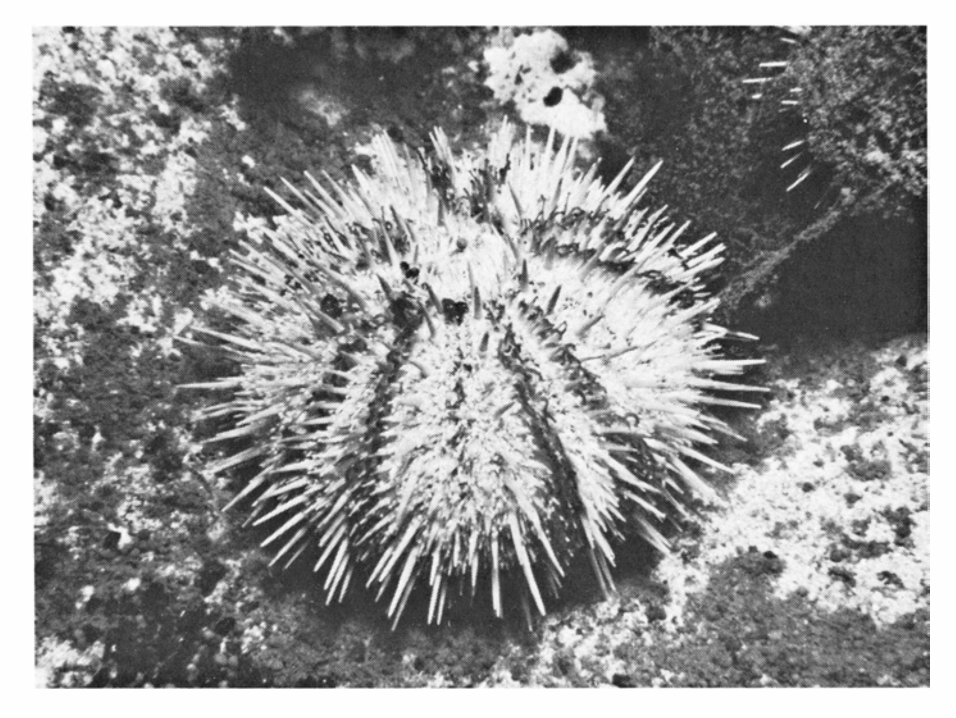
El desarrollo de un complicado esqueleto cristalino en el interior del embrión transparente del erizo de mar proporciona un modelo que explica la conformación de tejidos mineralizados duros tales como conchas y huesos

Shinya Inoué y Kayo Okazaki

El desarrollo embrionario a través de la división sucesiva, redistribución y diferenciación de las células a partir de un único huevo fertilizado constituye una cuestión básica en biología. Una parte de esta cuestión es saber cómo el embrión desarrolla, mediante las sustancias inorgánicas, estructuras cristalinas duras tales como los huesos, los dientes y la concha. Estos materiales, que pueden caracterizarse como biocristales, están formados por una distribución regular tridimensional de calcio, sílice, fosfato, carbonato o alguna combinación de estas sustancias; no se los puede distinguir químicamente de los cristales que se hallan en el mundo inanimado. Son muy notables, porque no se presentan en las formas geométricas simples de los cristales no biológicos, sino en formas intricadamente esculpidas que varían según los organismos vivos que los producen.

Los biocristales se encuentran en todo el reino animal, desde los esqueletos de sílice de las esponjas a los cristales estratificados de fosfato cálcico.

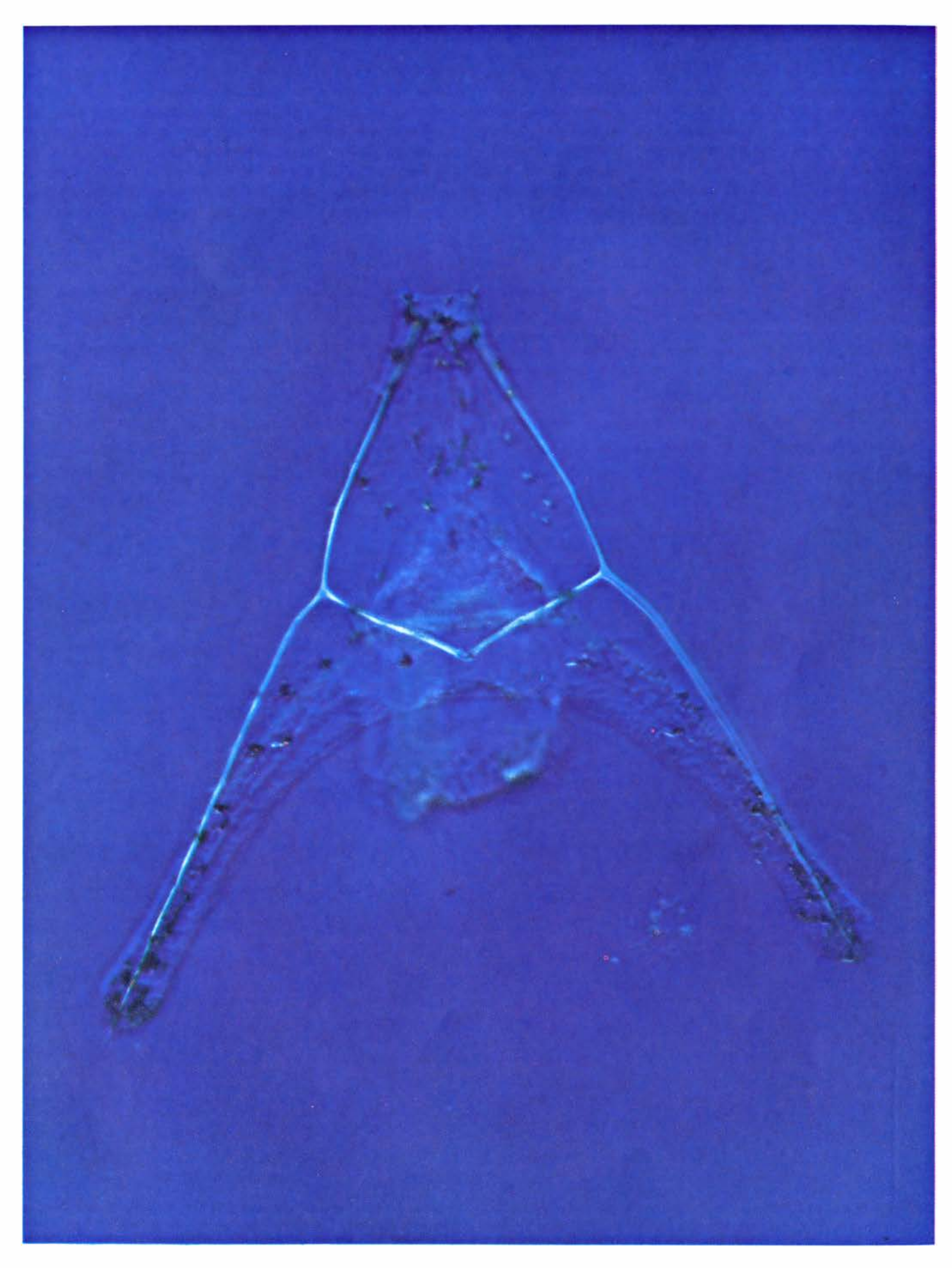
El erizo de mar, organismo marino cubierto de espinas que vive en aguas poco profundas de los arrecifes coralinos o en las lagunas originadas por las mareas, constituye un organismo excelente para la investigación de los biocristales. En su forma embrional produce dos espículas o biocristales de forma de estrella, que con el tiempo se unen entre sí para formar el esqueleto de la forma larvaria, provista de cuatro cuernos. Además, como el embrión es transparente, permite una visión clara de las células vivas en el proceso de transformación de los elementos minerales en estructuras cristalinas.



EL ERIZO DE MAR ADULTO es un habitante corriente de las lagunas de marea marina y de las aguas someras de los arrecifes coralinos. Al madurar produce millones de huevos o esperma.

El crecimiento de las espículas en el proceso de desarrollo del embrión del erizo de mar se ha descrito muy detalladamente a lo largo de los últimos cincuenta años. Al alcanzar la madurez, el erizo de mar adulto nutre millones de células germinales maduras (huevos o esperma) en el interior de su caparazón espinoso. Ya sea de manera natural o por una provocación artificial (tal como un estímulo eléctrico o la inyección de cloruro potásico) el organismo expulsa las células germinales a través de 5 poros microscópicos. En el agua de mar circundante, el esperma y los huevos liberados por los erizos macho y hembra se unen, y los huevos fertilizados empiezan a dividirse. Primero, los huevos se parten tres veces simétricamente, formando dos pisos de cuatro células de igual tamaño montadas unas sobre otras. En la cuarta división, las células se dividen de manera desigual, originando ocho células de tamaño intermedio, cuatro células grandes y otras cuatro menores llamadas micrómeros. La mayoría de las células continúa dividiéndose hasta que se forma la blástula: una esfera hueca formada por una capa única de células. Unas cuantas de estas células penetran en el interior de la blástula mientras siguen manteniendo contacto con sus vecinas, formando gástrula.

Para entonces, los descendientes de los micrómeros, las cuatro células pequeñas que se originaron en la cuarta división, han emigrado al interior de la cavidad de la gástrula. Se las denomina ahora células del mesénquima primario, y con la ayuda de pseudópodos –extensiones contráctiles largas– se asocian en dos grupos, uno de cada lado del intestino primario. Allí se unen los pseudópodos y en su interior aparece un pequeño gránulo calcáreo. A continuación, las células del mesénquima primario, conectadas

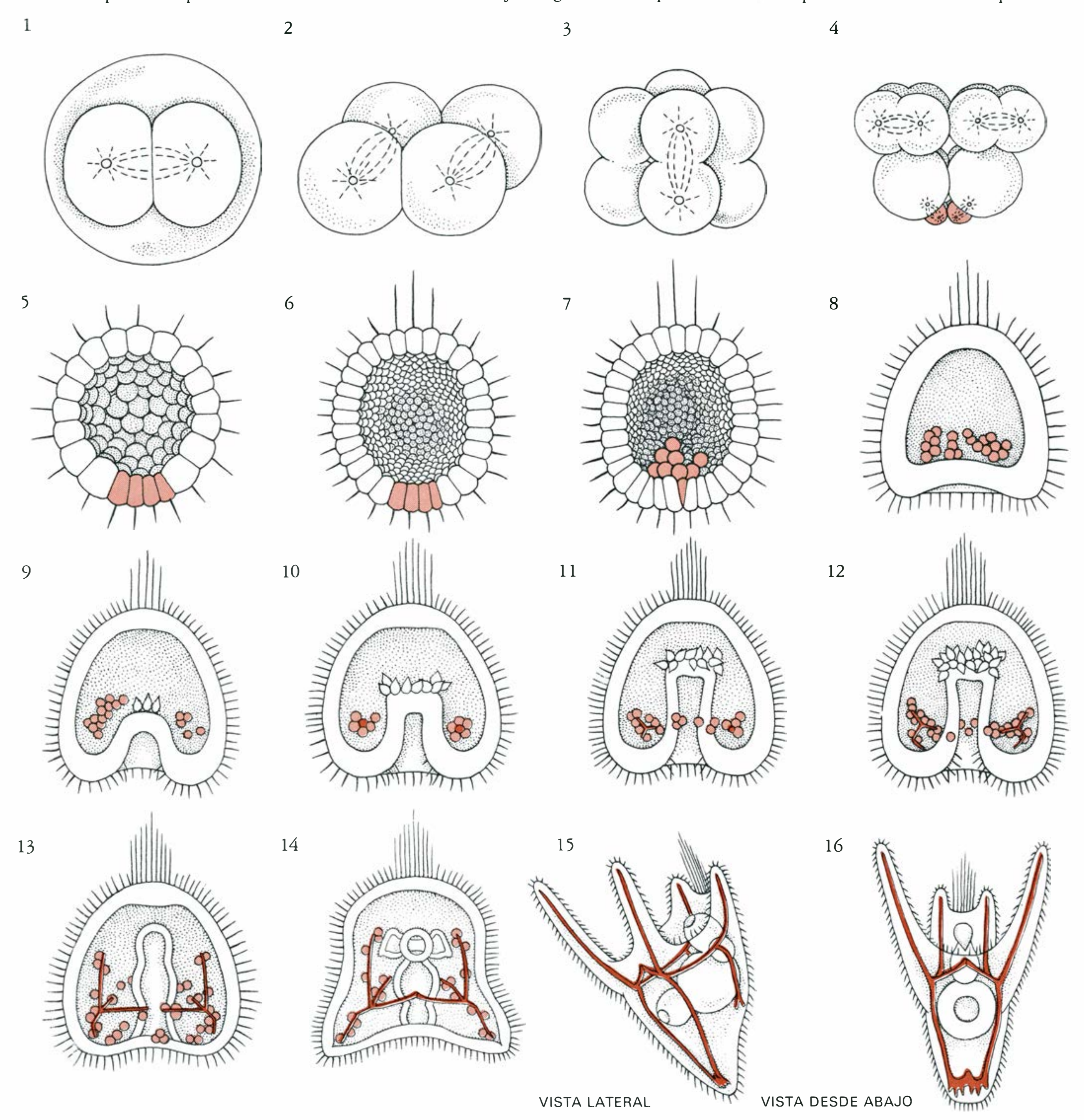


EL ESQUELETO CRISTALINO brilla dentro de una larva de erizo de mar en esta micrografía obtenida por uno de los autores (Inoué) usando microscopía de interferencia diferencial y de luz polarizada. Se ven dos de los cuatro cuernos de la larva. Ampliación: 640 diámetros.

aún por sus pseudópodos, se organizan en tres cables radiales separados 120° entre sí. A medida que el embrión continúa su desarrollo, los cables pseudopódicos guían el crecimiento del gránulo calcáreo hasta formar una espícula trirradiada, que es un cristal en forma de de estrella de tres puntas.

Más adelante, las cadenas de células del mesénquima adquieren otras configuraciones, emigrando cada cadena en una dirección diferente. Los cables pseudopódicos esculpen la espícula plana trirradiada, dando origen a una forma tridimensional que se denomina espícula esquelética. El proceso se inicia con un leve codo en ángulo recto en uno de los brazos de la espícula trirradiada. La porción arqueada se alarga y se transforma en el eje largo de la espícula esquelética, de la cual se proyectan una o varias varillas, según la especie de que se trate. Con el tiempo, un par de espículas esqueléticas unidas entre sí, que forman algo así como una silla plegable, proporcionan un esqueleto interior a la larva tetracorne del erizo de mar. La forma de la espícula larvaria es específica de cada especie del erizo de mar.

La espícula larvaria se compone de



DESARROLLO DEL ERIZO DE MAR a partir del huevo fertilizado hasta llegar a la forma larvaria. Una vez fertilizado, el huevo se divide tres veces *(1-3)*. Las ocho células que resultan se dividen de manera desigual *(4)*, originando ocho células de tamaño intermedio, cuatro células grandes y cuatro menores llamadas micrómeros *(color claro)*. Las células en división forman una esfera hueca o blástula *(5,6)*. Las descendientes de los micrómeros, las células del mesénquima primario, emigran a la cavidad blastular *(7, 8)*. Un doblez hacia el interior de la pared blastular forma el aparato digestivo primitivo del embrión, que ahora se denomina gástrula; las células del mesénquima se asocian a cada lado del pliegue *(9)*. Ahí los pseudópodos, o extensiones contráctiles, de las células mesenquimáticas se unen para formar cadenas radiales o cables. En el vértice de los cables aparece un gránulo calcáreo *(color oscuro)* *(10)*. Los dos gránulos de cada embrión forman espículas trirradiales planas *(11-12)* y luego espículas esqueléticas tridimensionales complementarias. Por último éstas se funden y forman el esqueleto de la larva *(13-16)*.

carbonato cálcico (CO_3Ca) con una pequeña cantidad de magnesio (5 por ciento) y matriz orgánica (1 por ciento). El carbonato cálcico, que es también el constituyente principal del coral, la caliza y el mármol, cristaliza en formas diferentes, pero todas ellas exhiben la propiedad óptica de la birrefringencia: dividen las ondas luminosas en dos componentes que vibran en planos mutuamente perpendiculares. Cuando un cristal de este tipo se coloca en un haz de luz polarizada (esto es, luz cuyas ondas vibran sólo en un plano) y se mira a través de un filtro polarizante, se observa un fenómeno característico. Cuando el plano de polarización se orienta en una dirección que no es ni paralela ni perpendicular a los planos de los átomos del cristal, el cristal se ilumina intensamente.

Hay pocas cosas tan llamativas como la observación de las larvas del erizo de mar hecha en un microscopio polarizante. Cuando los ejes de los polarizadores de dicho microscopio se cruzan de manera que se oscurece el fondo, las dos espículas esquéleticas del interior de cada larva que está nadando se iluminan y se oscurecen, alternativamente, como las estrellas. En cada rotación de 360° de una larva, cada espícula brilla cuatro veces en orientaciones exactamente perpendiculares entre sí.

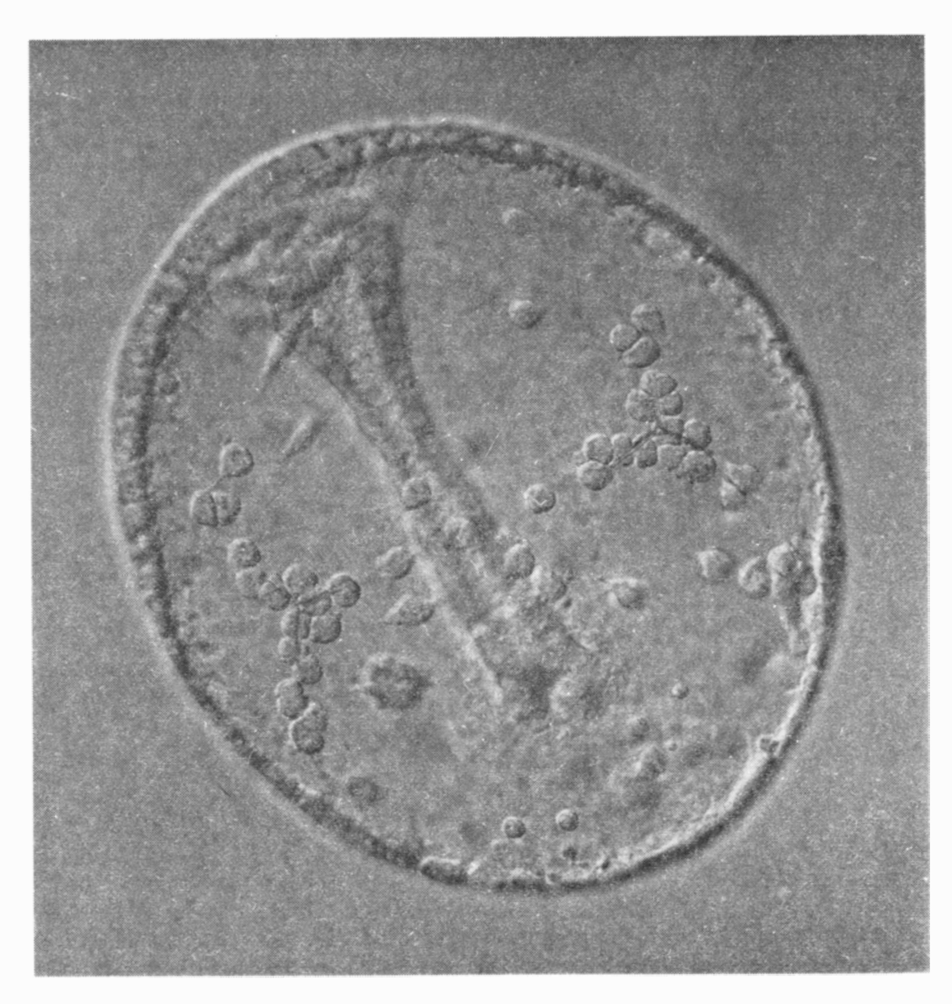
La birrefringencia de la espícula de la larva del erizo de mar es fuerte en comparación con la de otros materiales birrefringentes: el índice de refracción de la espícula es aproximadamente un 12 por ciento mayor cuando el plano de polarización de la luz polarizada es perpendicular al eje largo de la espícula que cuando el plano es paralelo a dicho eje. Este comportamiento, junto con la información que se obtiene por difracción de rayos-X, sugiere que el carbonato cálcico de la espícula cristaliza en forma de calcita. La estructura de la calcita consiste en planos de iones calcio, que tienen dos cargas positivas, alternando con planos de grupos carbonato (CO_3), que tienen dos cargas negativas. Los planos de calcio y carbonato son perpendiculares al eje cristalográfico del cristal y de esta manera definen dicho eje. En la calcita, las fuerzas de enlace son débiles perpendicularmente a los planos de las caras de un romboedro, de tal manera que los cristales no biológicos de calcita tienen una forma casi perfecta de romboedro.

La birrefringencia que exhibe la espícula esquelética de la larva del erizo de mar nos indica que cada espícula, a pesar de su forma externa compleja, se comporta ópticamente como si hubiera sido tallada de un cristal mayor de calcita no biológico. Este cristal mayor estaría orientado de tal manera que su eje cristalográfico coincidiese con la longitud de la espícula. Por lo tanto, la espícula esquelética de la larva del erizo de mar aparentemente combina la forma interna de un cristal único de calcita con la forma externa característica de la especie.

Esta estructura nos fascinó por vez primera hace años, cuando estábamos trabajando con un microscopio polarizante de construcción casera en el laboratorio de Katsuman Dan, en la estación de biología marina Misaki, al sur de Tokio. Hace dos veranos reanudamos nuestra colaboración en el laboratorio de biología marina de Woods Hole, Massachusetts, donde uno de nosotros (Okazaki) tenía una beca. Kent McDonald, un estudiante graduado de la Universidad de Colorado, trabajaba con el microscopio electrónico de barrido del laboratorio. Pensamos que seguramente dicho instrumento resultaría ideal para observar la estructura detallada de la espícula de la larva del erizo de mar con el fin de ampliar el conocimiento que habíamos obtenido con el microscopio polarizante. Para ello, aislamos las espículas trirradiadas y las espículas esqueléticas colocando los embriones del erizo de mar en agua dulce que hace que las células del embrión absorban agua por ósmosis y revienten. Las espículas se lavan bien con hidróxido sódico diluido, o Clorox (los mismos agentes que se emplean para limpiar los caparazones de erizo de mar adulto que se encuentran en las tiendas de recuerdos).

El microscopio electrónico de barrido mostró de manera espléndida las distintas espículas. La sencilla elegancia de las espículas trirradiadas y la belleza escultural de las espículas esqueléticas eran maravillosas. Pero había una cosa extraña. Independientemente de la ampliación que se usara –y con los mayores aumentos podíamos resolver estructuras de 100 angstroms (10^{-5} mm) de diámetro– no se percibía signo alguno de textura cristalina en la superficie de la espícula aislada.

Intentamos corroer las espículas con ácidos débiles y con un reactivo de gran



UNAS CADENAS DE CELULAS MESENQUIMATICAS guían la formación de las dos espículas trirradiales en la cavidad del embrión del erizo de mar. Ampliación: 500 diámetros.

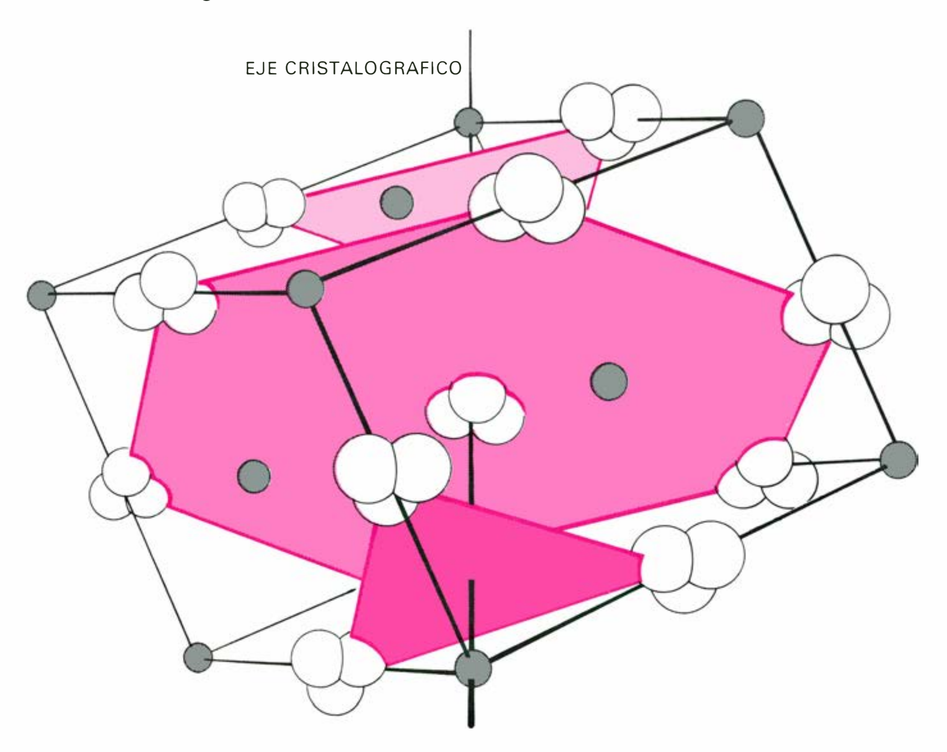
afinidad para con el calcio. La superficie se erosionó, pero no pudimos encontrar traza alguna de forma cristalina geométrica, tal como se encuentra en los cristales no biológicos de calcita atacados por ácidos débiles. Por otra parte, la calcita no biológica tiende a exfoliarse según ciertos planos cristalográficos y fácilmente presenta caras cristalinas que forman, entre sí, ángulos romboédricos característicos. Sin embargo, cuando quebrábamos espículas esqueléticas no podíamos encontrar caras de exfoliación con el microscopio electrónico de barrido. Por el contrario, las espículas parecían romperse según caras lisas e irregulares, tal como lo hacen los materiales no cristalinos como el vidrio. En resumen, aunque las espículas esqueléticas se comportan ópticamente como si fueran cristales únicos, el microscopio electrónico de barrido no pudo revelar ningún signo de cristalinidad.

Intentando resolver esta paradoja, decidimos ver si podíamos usar espículas aisladas como semillas sobre las cuales pudieran crecer cristales de calcita. Si ésto podía hacerse, las moléculas de carbonato cálcico en el cristal en crecimiento se alinearían con moléculas similares de la espícula biocristalina, revelando de esta manera la estructura cristalina subyacente. Introdujimos, por lo tanto, espículas aisladas en una solución saturada de carbonato cálcico; para lo cual las sumergimos primero en una solución acuosa de bicarbonato sódico, a la que se añadió después una solución de cloruro cálcico.

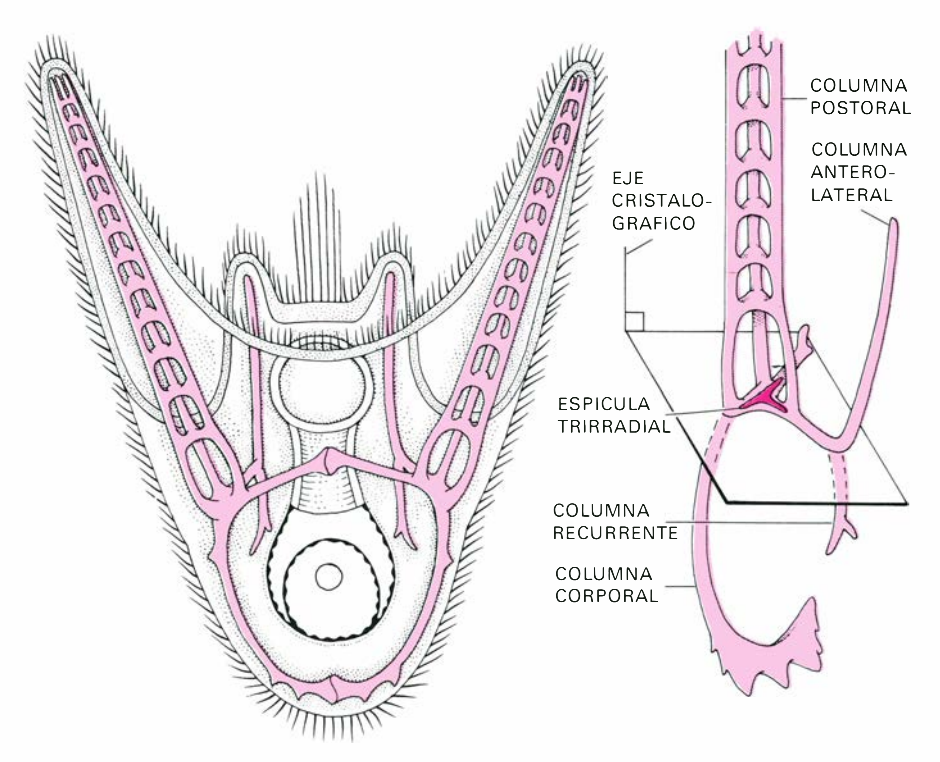
Cuando observamos las espículas semilla en el microscopio polarizante, aparecieron más gruesas y brillantes en pocos minutos. Las aristas de cada espícula se hicieron aserradas y, a pesar de todo, la espícula agrandada seguía oscureciéndose cuando su orientación era la misma que cuando se oscurecía antes del tratamiento. Parecía, por tanto, que el carbonato cálcico se estaba depositando en la superficie de la espícula con las moléculas alineadas en el mismo orden que las moléculas de la espícula.

El microscopio electrónico de barrido confirmó inmediatamente esta interpretación. Tanto las espículas trirradiadas como las esqueléticas estaban "decoradas" con diminutos cristales romboédricos, como un conjunto apilado de baldosas cristalinas. Fuera cual fuese la parte de la espícula que estuviéramos observando, todas las caras cristalinas correspondientes eran paralelas entre sí, independientemente del tipo de ramificación, esculturación, arqueo o alargamiento de la espícula. El hecho de que el ángulo entre las caras adyacentes de los cristales calcáreos fuese el característico de la calcita se comprobó al observar un par estereoscópico de micrografías electrónicas de un cristal decorador puesto al lado de un romboedro de calcita no biológica.

Resultaba, pues, que habíamos amplificado con éxito la distribución cristalográfica de las moléculas de carbonato cálcico de la espícula. Puesto que todas las superficies de los cristales decoradores

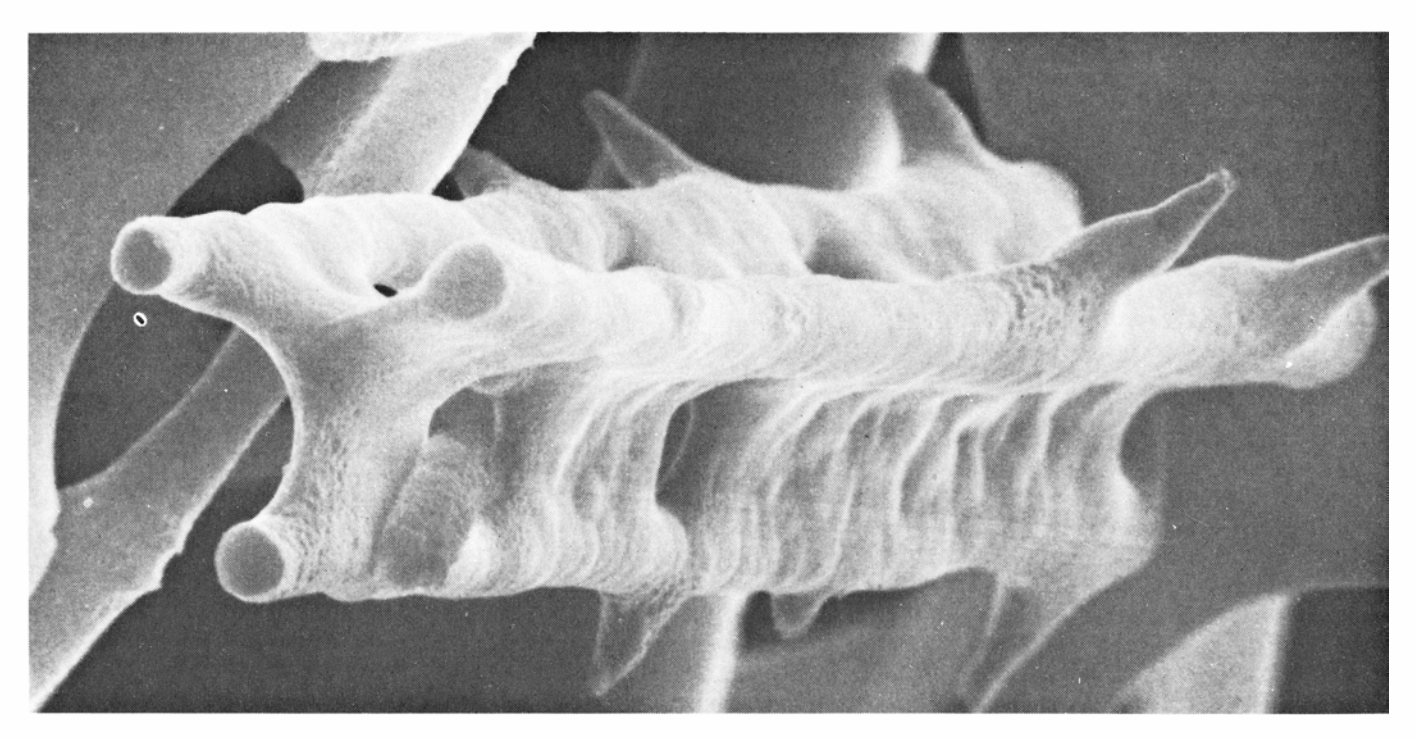


LA ESTRUCTURA DE LA CALCITA consiste en planos alternantes de iones calcio *(gris)* y grupos carbonato *(blanco)* perpendiculares a un eje cristalográfico vertical. Las fuerzas de enlace en el cristal son débiles a lo largo de las caras de un romboedro, originando así su forma característica.

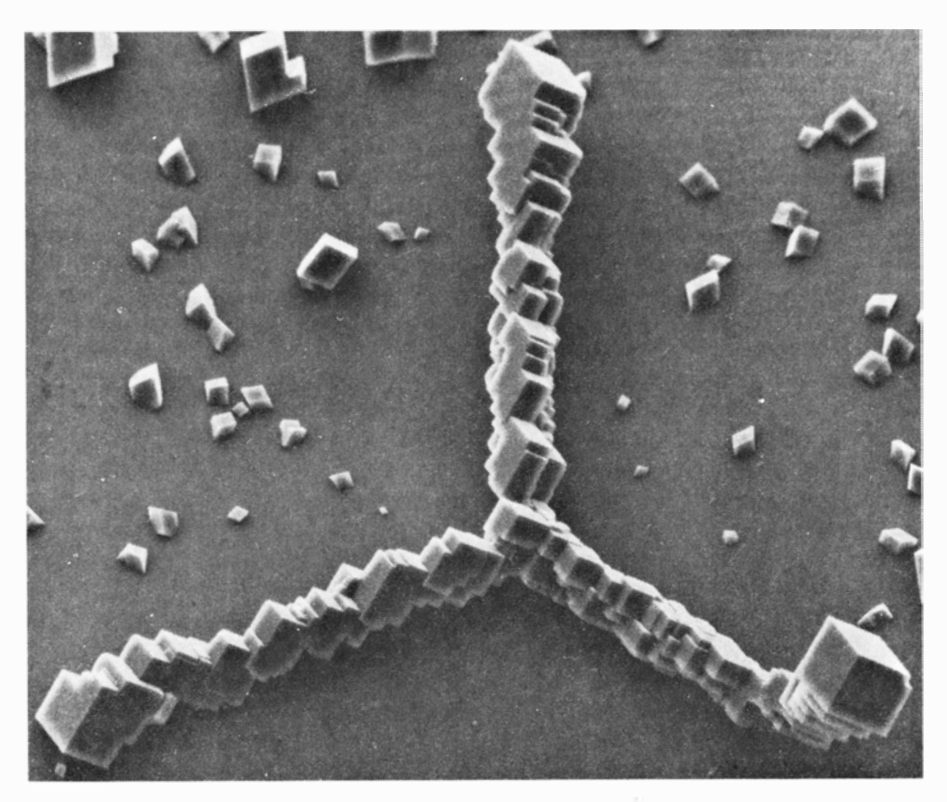


UN PAR DE ESPICULAS ESQUELETICAS *(color claro)* unidas entre sí forman el soporte de la larva de cuatro cuernos del erizo de mar *Arbacia punctulata.* El esquema de la derecha indica cómo las columnas de cada espícula esquelética se desarrollan a partir de los brazos de una espícula trirradial plana *(color oscuro).* Cada espícula esquelética, a pesar de su forma compleja, se comporta ópticamente como si hubiera sido esculpida a partir de un cristal de calcita único.

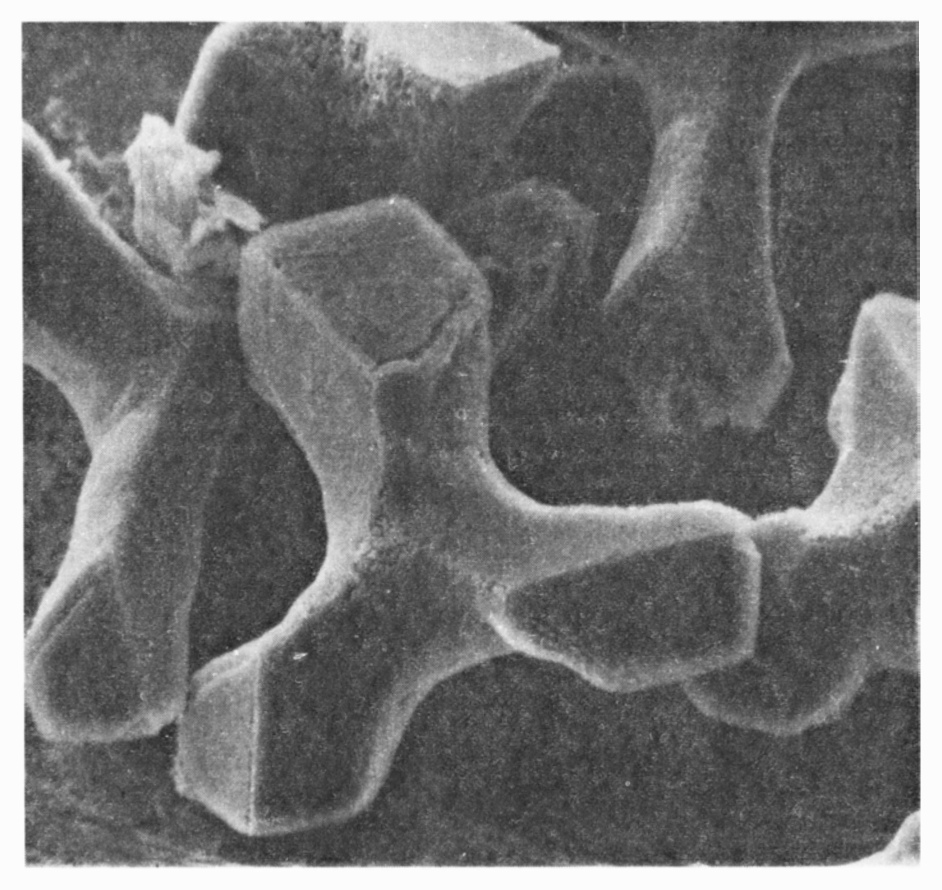
ESPICULAS TRIRRADIALES extraídas de embriones de erizo de mar. Micrografía electrónica de barrido con una ampliación de 4.200 diámetros. Un brazo de cada espícula trirradial ha iniciado un codo en ángulo recto que formará la columna principal de la espícula esquelética.



DETALLE DE ESPICULA ESQUELETICA de *Arbacia,* mostrando la complicada superficie de la columna postoral; micrografía electrónica de barrido con ampliación de 8.600 diámetros. Se observa la ausencia de caras cristalinas en la superficie externa lisa de la espícula y en las tres superficies de fractura a la izquierda que, en cambio, son lisas y cóncavas como en un material no cristalino como el vidrio.



UNA BALDOSA DE CRISTAL DE CALCITA "decora" completamente una espícula trirradial que se aisló y se empleó como semilla para crecimiento cristalino. La espícula se incubó en una solución de cloruro cálcico y bicarbonato sódico. Los ejes de los cristales decoradores son exactamente paralelos entre sí, indicando que la distribución de moléculas de carbonato cálcico del biocristal que sirvieron de plantilla es idéntica a la del cristal único de calcita no biológico.



ESPICULAS TRIRRADIALES ANORMALES inducidas experimentalmente al hacer crecer embriones de erizo de mar en agua marina con deficiencia de magnesio y cálcio. Estas condiciones debilitan las células del mesénquima, que ya no pueden formar cables pseudopódicos normales. Por eso, las fuerzas inorgánicas que actúan en el cristal de calcita vencen a las fuerzas biológicas ejercidas por las células y aparecen caras cristalinas en los brazos de las espículas trirradiales.

eran paralelas entre sí, todas las moléculas de carbonato cálcico de la espícula que habían servido como molde o plantilla tenían también que estar orientadas exactamente paralelas entre sí. De hecho, los ejes cristalográficos de los cristales que decoraban una espícula única estaban orientados siempre de forma paralela a los ejes cristalográficos de la propia espícula, y lo mismo sucedía a cada una de las espículas derecha e izquierda de un par fundido, de manera que, en la unión entre las dos espículas, los dos conjuntos de cristales se unían con sus ejes cristalográficos inclinados uno respecto del otro, con el mismo ángulo que mantenían entre sí las espículas derecha e izquierda.

Las moléculas de carbonato cálcico de cada espícula están por lo tanto distribuidas de forma regular, como lo hacen en un cristal único de calcita no biológico. A pesar de ello, el biocristal, en vez de aparecer con la forma externa que presenta la calcita no biológica, adquiere una forma complicada característica de cada especie de erizo de mar. ¿Cómo realiza esta hazaña el embrión del erizo?

Según quedó dicho, la formación de cada espícula embrional que se desarrolla viene guiada por tres cables pseudopódicos radiales, o cadenas de células mesenquimatosas. Estas células impiden el crecimiento incluso de la más mínima cara cristalina en la espícula, evitando así la formación de puntos mecánicamente débiles y de superficies abrasivas que pudieran irritar a las células de los tejidos vecinos. En una serie de elegantes experimentos, Sven Hörstadius, de la Universidad de Estocolmo, mostró que el núcleo de la célula micrómera (la precursora de las células del mesénquima) es justamente el que lleva los mensajes hereditarios que instruyen a las células en cuanto se refiere a la manera y momento de crecer el biocristal, para generar así la conformación singular de la especie. Hasta hoy sigue siendo un misterio intrigante la manera en que las células del mesénquima descifran estos mensajes y conocen sus actuaciones individuales en la formación colectiva del biocristal.

Recientemente, uno de nosotros (Okazaki) ha logrado aislar micrómeros de los embriones del erizo de mar y hacerlos crecer en cultivo de tejidos. Las veinte y tantas células mesenquimáticas descendientes de cada micrómero podían fabricar espículas con la forma característica de la especie. Esto confirma que los micrómeros solos pueden expresar información genética para dirigir el crecimiento de las espículas esqueléticas.

¿Qué pasaría si el embrión no pudiera alinear los cables pseudopódicos de manera normal? Hace unos cuantos años, uno de nosotros (Okazaki) depositó embriones con espículas trirradiadas en agua de mar con una concentración de iones calcio menor que la normal. Las células del mesénquima formaron entonces cables pseudopódicos menos extensos, y las puntas normalmente finas de las espículas se hicieron bulbosas. Si las concentraciones de calcio y magnesio en el agua del mar se reducían a la décima parte de la concentración normal, las puntas bulbosas se hacían angulosas, como si allí estuvieran creciendo cristales de calcita.

Observando esta espícula en tres dimensiones junto a un romboedro de calcita no biológica con orientación similar (gracias a pares estereoscópicos de micrografías electrónicas de barrido) vimos que cada cara de las tres puntas de las espículas trirradiadas era exactamente paralela a las caras de un único romboedro de calcita.

En un embrión privado de calcio y magnesio, por tanto, las caras de calcita aparecen en la espícula en crecimiento. Las fuerzas que dirigen el crecimiento de calcita no biológica han dominado sobre el mecanismo controlador de las células y han formado caras cristalinas que son las que corresponden a las fuerzas de enlace molecular en la calcita y no a las del ordenamiento que se observa normalmente en el biocristal. Para expresarlo de otra manera, las células no pueden manifestar la estructura codificada en sus genes, y domina la tendencia del cristal no biológico a formar sus propias caras características.

En resumen, cuando los organismos vivos están formando los componentes minerales de estructura dura, parece que se aprovechan de la tendencia cristalizante de sustancias tales como el carbonato cálcico y el fosfato cálcico, utilizando las fuerzas que mantienen las moléculas ordenadas regularmente en un cristal único; a pesar de ello, regulan el crecimiento del cristal hasta obtener el tamaño y la forma apropiados para el organismo, según el tiempo y el lugar. El biocristal expone vívidamente el juego entre fuerzas moleculares inanimadas, que tienden a una configuración de energía mínima, y la capacidad de organización de las células vivas. La vida sigue por necesidad principios inanimados, pero en vez de hacerlo por la fuerza bruta parece canalizar el flujo de energía, guiando el ordenamiento de la materia hacia formas cada vez más complejas y termodinámicamente improbables.

Los acompañantes de las estrellas de tipo solar

Muchas estrellas tienen a otras por acompañantes, ¿poseen planetas algunas de ellas? La búsqueda espectrográfica de acompañantes en 123 estrellas próximas aporta la clave de la frecuencia de dichos sistemas planetarios

Helmut A. Abt

El sol se acompaña en su viaje a través del espacio por un séquito de nueve planetas, 32 satélites naturales que giran en torno a éstos, cientos de cometas y miles de asteroides. ¿Tienen otras estrellas parecidas formaciones de acompañantes? No existe, en la actualidad, ninguna vía directa para responder a ello, pues aunque semejante colección de cuerpos estuviera asociada a una estrella cercana, los telescopios más potentes del mundo no podrían detectarlos. La débil luz reflejada por los acompañantes se perdería en el brillante resplandor de la estrella central.

Algún día, tal vez, cuando grandes telescopios instalados en el espacio permitan a los astrónomos ver las estrellas sin el estorbo de las distorsiones de la atmósfera terrestre, será posible resolver el más brillante de los acompañantes planetarios en torno a otra estrella. Con instrumentos auxiliares más sensibles, los telescopios terrestres pueden llegar incluso a determinar si una estrella está siendo perturbada por algún gran planeta que gire en torno a la misma. Hay algunas pruebas respecto de que ciertas estrellas próximas, en particular la estrella de Barnard, pueden estar escoltadas por esos "acompañantes oscuros" no vistos, pero los datos disponibles no son concluyentes.

Por otra parte, son muchas las estrellas de las que se sabe que tienen otras por acompañantes. En unos casos, los miembros de la pareja estelar son idénticos en brillo. En otros, la estrella secundaria brilla muy poco. El sol no parece tener ninguna estrella acompañante, aunque no está totalmente excluida la posibilidad de que exista alguna que lo sea, muy apagada y alejada de los confines del sistema solar. ¿Cabe la posibilidad de que los sistemas estelares múltiples se formen según una variedad de combinaciones, que iría desde parejas de casi el mismo brillo, pasando por las de brillos muy diferentes, hasta estrellas solitarias con planetas?

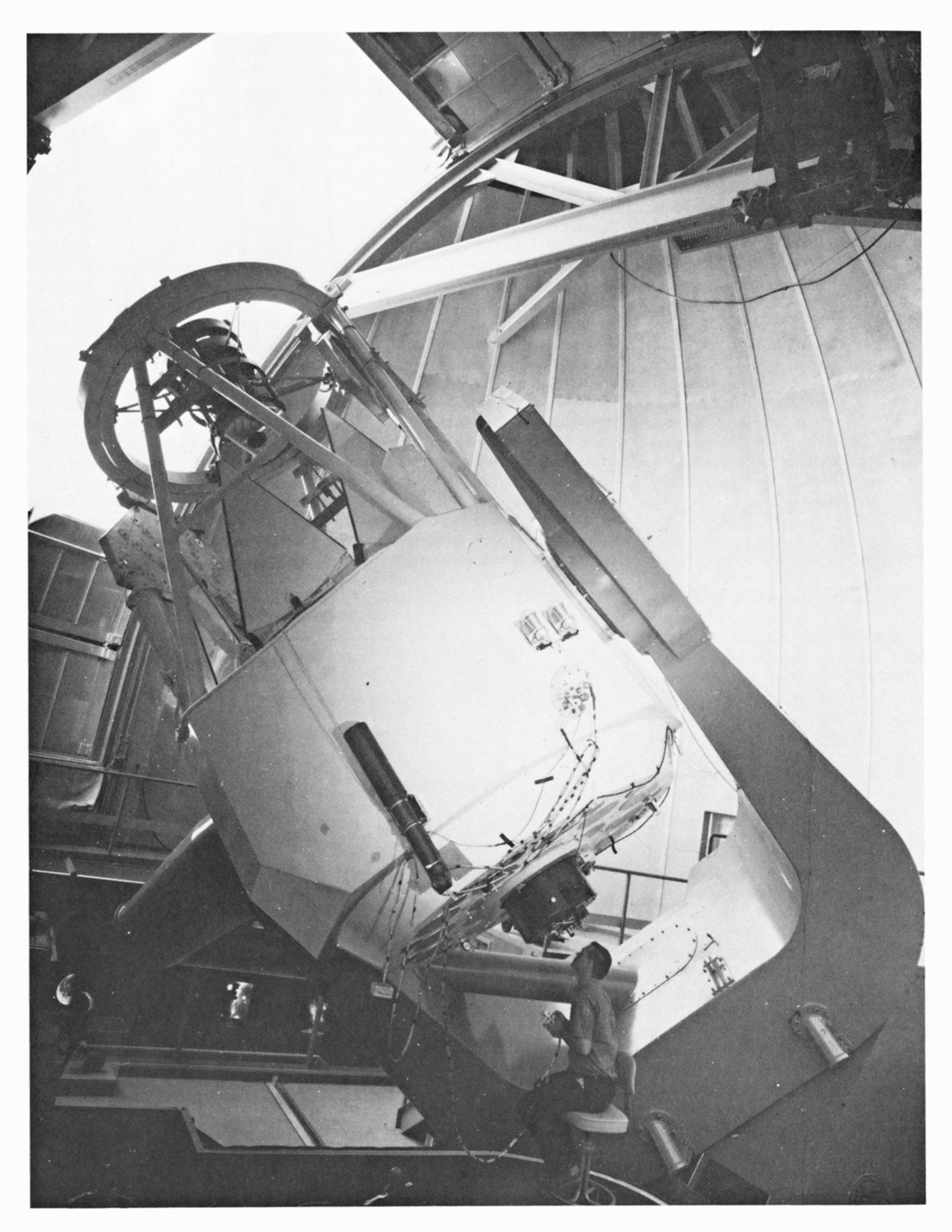
En 1960, Frank D. Drake, que trabajaba a la sazón en el National Radio Astronomy Observatory de Green Bank, Estados Unidos, ideó un experimento llamado Proyecto Ozma, destinado a detectar otras civilizaciones mediante la escucha de señales de radio inteligibles procedentes de estrellas cercanas. Para sus primeros intentos, Drake seleccionó aquellas de las que no se sabía que tuvieran estrellas acompañantes conocidas, ya que le parecían que serian también las dotadas de mayores probabilidades de tener planetas. ¿Era éste un planteamiento razonable?

En otro esfuerzo por responder a ese y otros interrogantes, mi colega Saul Levy y yo emprendimos recientemente otro tipo de investigación en el Kitt Peak National Observatory, en Arizona. Estudiamos los espectros de un gran número de estrellas relativamente brillantes, buscando claves sobre la posible existencia de acompañantes no detectados. Al hacerlo, obtuvimos algunas informaciones nuevas importantes, sobre la frecuencia con que se presentan distintas combinaciones de estrellas múltiples y sobre las probabilidades de encontrar otros sistemas planetarios.

Como estrella, el sol no difiere grandemente, a buen seguro, de otras muchas estrellas del cielo. Si bien es cierto que pueden verse más rasgos –manchas y protuberancias– en el sol que en ninguna otra estrella, ello se debe a que aquél es la única estrella cuya superficie puede resolverse fácilmente desde la tierra. Las demás estrellas aparecerían como alfilerazos de luz, incluso a través de los mayores telescopios, si no fuera por el hecho de que la turbulenta atmósfera de la tierra difumina esos alfilerazos y forma discos aparentes. Hay diversos métodos, muy complicados, que ponen de manifiesto rasgos globales de las superficies de otras estrellas; por ejemplo, el estudio de algunas cuando son gradualmente eclipsadas por otras. Sin embargo, parece razonable, en general, suponer que hay millones de estrellas esencialmente idénticas al sol. Las estrellas que tienen aproximadamente la misma temperatura superficial (6000 grados Kelvin), el mismo diámetro (1,4 millones de kilómetros) y la misma historia evolutiva que el sol, se llaman, por lo mismo, estrellas de tipo solar.

Fundamentalmente, se citan cuatro maneras de determinar si una estrella es miembro de un sistema múltiple de ellas. La vía más directa es observar (a través de un telescopio) dos estrellas que se hallen juntas y llevar un registro de sus posiciones relativas, para determinar si giran gradualmente en torno a un centro de gravedad común. Una vuelta completa puede durar décadas o siglos; ahora bien, con mediciones precisas cabe frecuentemente mostrar, durante una décima de revolución, que las estrellas están siguiendo trayectorias elípticas, en vez de las trayectorias rectilíneas, que seguirían las estrellas que se cruzaran accidentalmente en el espacio. Las parejas estelares identificadas de este modo se llaman dobles visuales.

A veces, las estrellas de una pareja están tan separadas que una sola vuelta de una en torno a la otra demoraría miles o millones de años. Tal vez no sea posible observar movimiento orbital alguno a lo largo de una, o incluso de varias generaciones humanas. Tales estrellas pueden aparecer en las mismas posiciones relativas durante décadas, mientras que otras dos estrellas, la una cercana y la otra muy alejada, que se diría que están proyectadas muy cerca entre sí en el cielo, aparecerían deslizándose gradualmente cada una por su lado, ya que cada



TELESCOPIO REFLECTOR DE 2,1 METROS del Kitt Peak National Observatory, Arizona, utilizado por el autor y Saul Levy para determinar la frecuencia de los sistemas estelares múltiples entre las 123 estrellas de tipo solar observables a simple vista desde el hemisferio norte. Se obtuvo un total de 20 espectrogramas de cada estrella, con el espectrógrafo *coudé (véase el diagrama de las dos páginas siguientes).* Para un telescopio tan grande, el reflector de 2,1 m. resulta muy compacto. Ello se debe a la semidistancia focal (f/2,63)de su espejo primario.

estrella tiene, típicamente, su propio movimiento a través del espacio. Las parejas de estrellas, que mientras se desplazan se mantienen juntas, se llaman pares con movimiento propio común.

En algunos casos, la estrella acompañante puede mostrarse muy pálida, o hallarse demasiado cerca de la estrella más brillante, como para ser observada directamente; ahora bien, la repetición de mediciones precisas de la posición de la estrella más brillante puede revelar que su trayectoria sobre el cielo sigue una línea ondulada. Las estrellas tienden a moverse en línea recta a menos que sean perturbadas por un acompañante, por lo que una estrella que presente esta clase de movimiento puede ser miembro de una pareja. El sistema estelar que pertenezca a ese tipo se llama doble astrométrica. Se conocen pocos sistemas de éstos, porque suelen necesitarse centenares de mediciones de la posición de una estrella aislada para detectar una perturbación significativa.

Por último, hay algunas estrellas que aparecen como aisladas en un telescopio, pero cuyos espectrogramas ponen de manifiesto que son dobles. Si la luz procedente de una estrella se esparce según sus colores (o longitudes de onda), por medio de un prisma o de una red de difracción, se observa una estructura de rayas de absorción, es decir, de huecos en el alineamiento de colores. La estructura de rayas es característica de la temperatura y de la composición de la estrella. Cada raya es producida por un determinado elemento químico que absorbe esa específica longitud de onda de la luz. Las posiciones relativas de las rayas se conocen con gran precisión (seis o siete decimales), pero en la mayoría de las estrellas toda la estructura está desplazada hacia longitudes de onda más cortas o más largas, debido al movimiento de la estrella con respecto al observador (efecto Doppler). Si la estrella se acerca, las líneas están corridas hacia el extremo violeta del espectro visible; si está alejándose, las rayas están corridas hacia el extremo rojo. Un corrimiento global de este tipo hacia el rojo, conocido como el corrimiento cosmológico hacia el rojo, es característico de las galaxias lejanas y proporciona la prueba fundamental en favor de la teoría, ampliamente aceptada, de que el universo está en expansión.

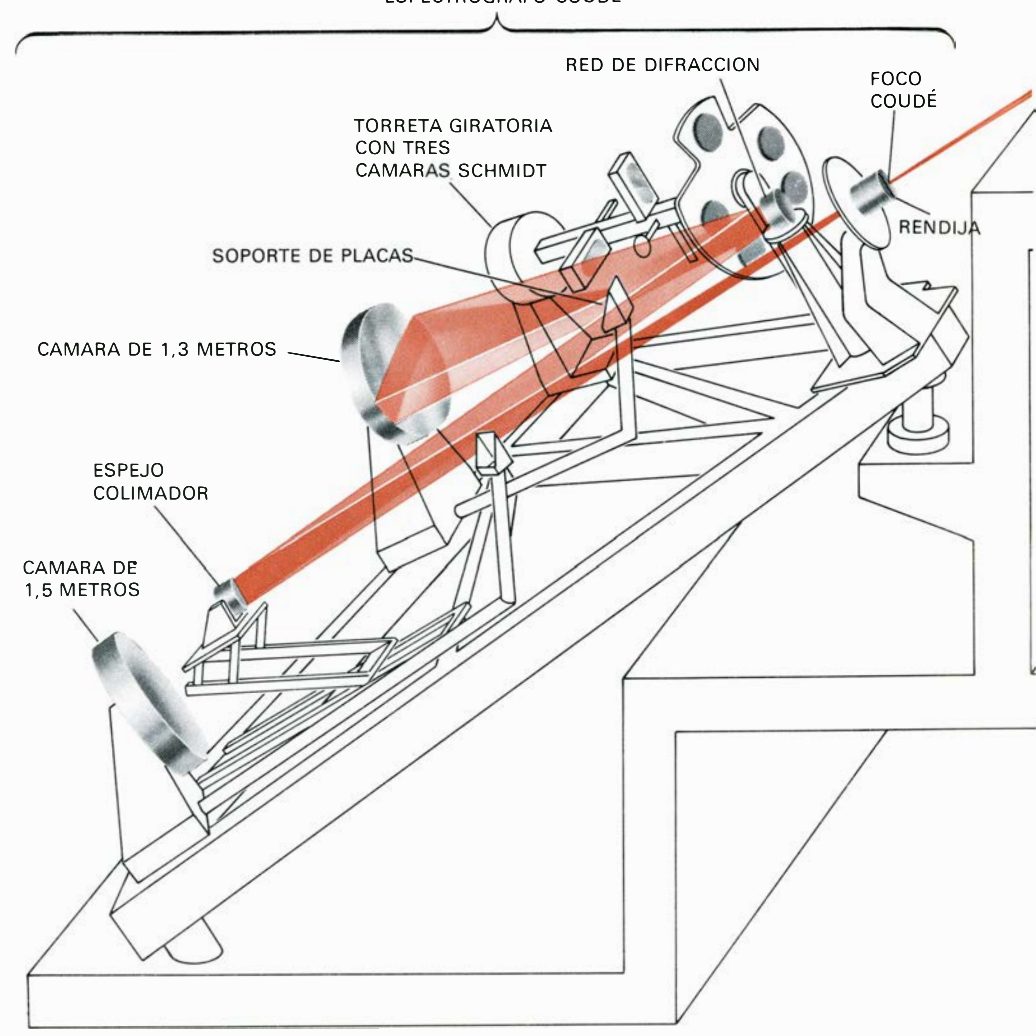
En algunas estrellas, los corrimientos Doppler varían con el tiempo. Unos días las rayas están desplazadas hacia el extremo violeta del espectro y otros lo están hacia el rojo. En la mayoría de estos casos, los corrimientos Doppler variables pueden atribuirse a una estrella que esté girando en una órbita en torno a un centro de gravedad común con un acompañante invisible. El acompañante se hace invisible por la sencilla razón de que su luz es mucho más débil que la luz de la estrella primaria. Un sistema tal se llama doble espectroscópica, toda vez que su duplicidad aparece únicamente con la ayuda del espectrógrafo.

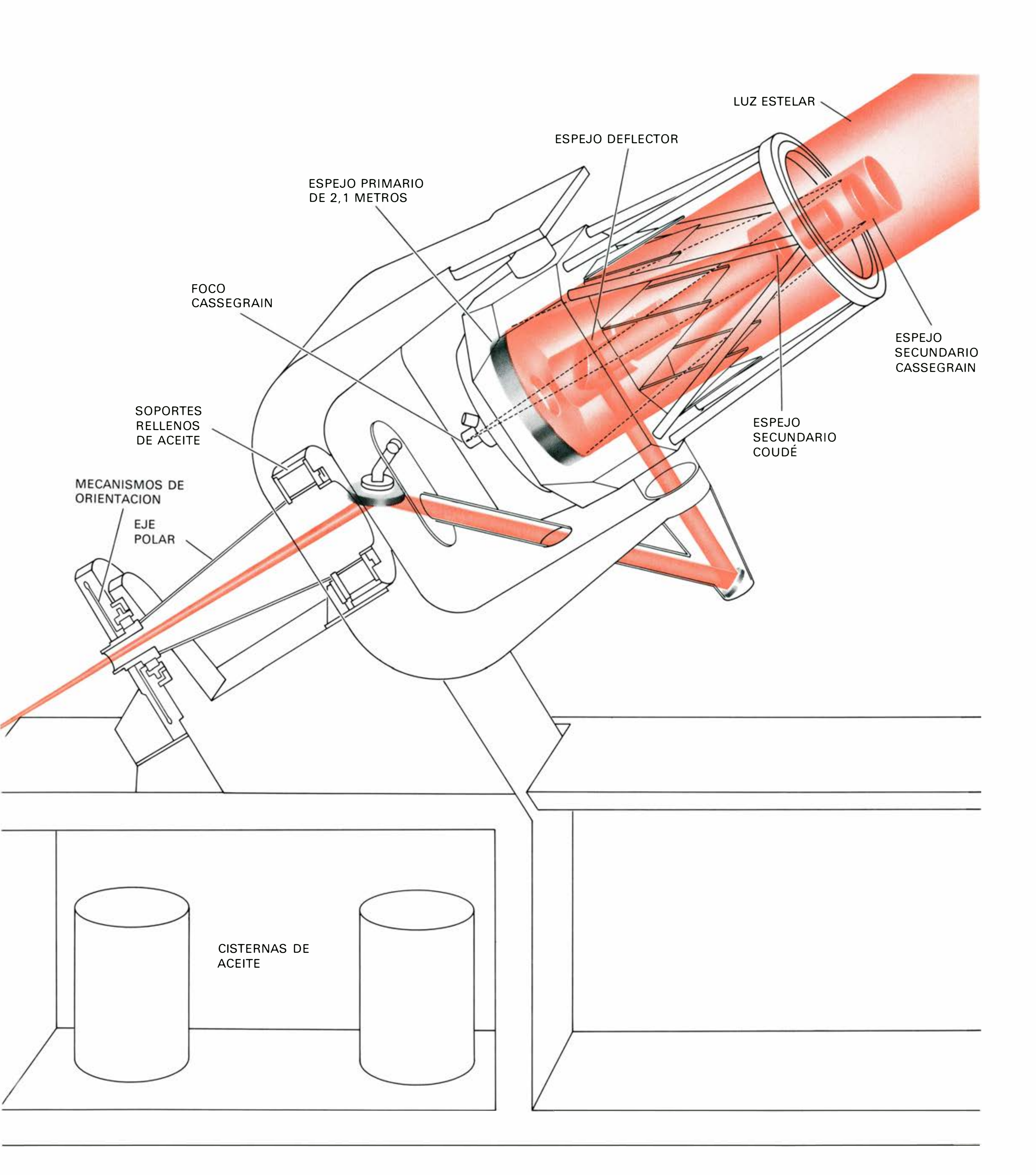
Si las dos componentes de una doble espectroscópica brillan con similar intensidad, los espectros de ambas estrellas se superponen en el espectrograma. Las rayas son dobles cuando una estrella se está acercando y la otra se está alejando; son simples cuando los movimientos orbitales forman un ángulo recto con la visual. A las parejas estelares de este tipo se las denomina dobles espectroscópicas de rayas dobles.

Los astrónomos han venido recogiendo datos sobre varios sistemas de estrellas dobles de entre las estrellas más brillantes de tipo solar, a lo largo de casi un siglo. Por los años 30 y, luego, en la década de los 60, se intentó determinar qué fracción de todas las estrellas observables a simple vista eran dobles, pero las estimaciones diferían substancialmente. Para unos, sólo el 18 por ciento de las estrellas de tipo solar eran dobles; para quienes militaban en el extremo opuesto, éstas llegaban al 54 por ciento. La diferencia provino de distintas interpretaciones sobre si los corrimientos Doppler de muchas estrellas eran realmente variables o si sólo reflejaban inexactitudes atribuibles a las mediciones.

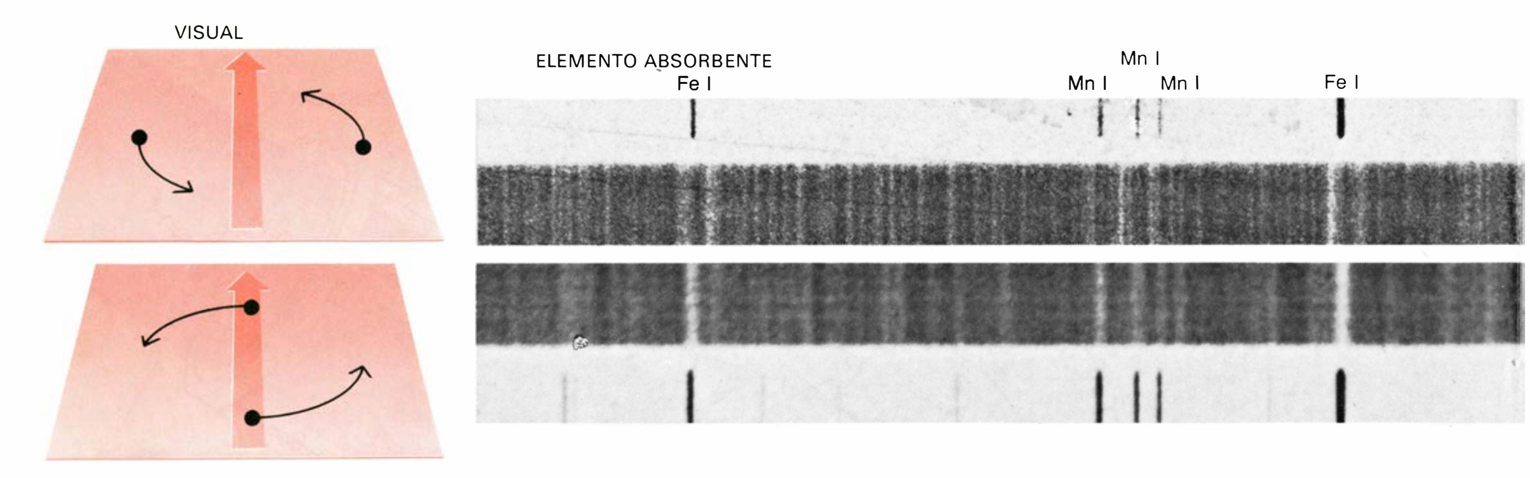
Hace algunos años, Levy y el autor decidieron determinar con mayor cuidado la frecuencia de las estrellas dobles y de otras multiplicidades entre las estrellas de tipo solar. Las razones del proyecto eran varias. Una estribaba en averiguar si, en el proceso de formación de las estrellas, la tendencia a constituir estrellas aisladas era mayor que la tendencia a formar múltiples. Otra razón era saber algo, si se podía, sobre la probable incidencia de planetas fuera del sistema solar. Había más motivos para emprender el proyecto, pero como sucede tan a menudo en la ciencia, muchos de los resultados más interesantes llegan por sorpresa.

En colaboración con Levy, el autor





ALTERNATIVAS PARA EL CAMINO OPTICO disponibles en el reflector de 2,1 metros de Kitt Peak. En el montaje Cassegrain, un espejo convexo colocado en el extremo del tubo del telescopio refleja la luz que viene del espejo primario, de 2,1 metros, a través de un agujero practicado en el centro del espejo primario, hacia un foco situado justo por debajo del extremo inferior del tubo. Se tiene acceso al foco Cassegrain por medio de una plataforma móvil de acero. En el montaje *coudé*, se hace girar un anillo interno situado en el extremo superior del telescopio para sustituir el espejo Cassegrain por el espejo secundario *coudé*. Se suceden, luego, tres espejos planos adicionales que reflejan la luz estelar hacia un foco al sur de la montura del telescopio, donde está colocado el espectrógrafo *coudé*. El camino óptico en zigzag es la opción que evita el problema del acoplamiento del espectrógrafo, de 13,5 toneladas, al telescopio. El espectrógrafo consta de cinco grandes cámaras, cuyos diámetros van de 0,5 a 1,5 metros; éstas fotografían los espectros producidos como resultado de la dispersión de la luz estelar.



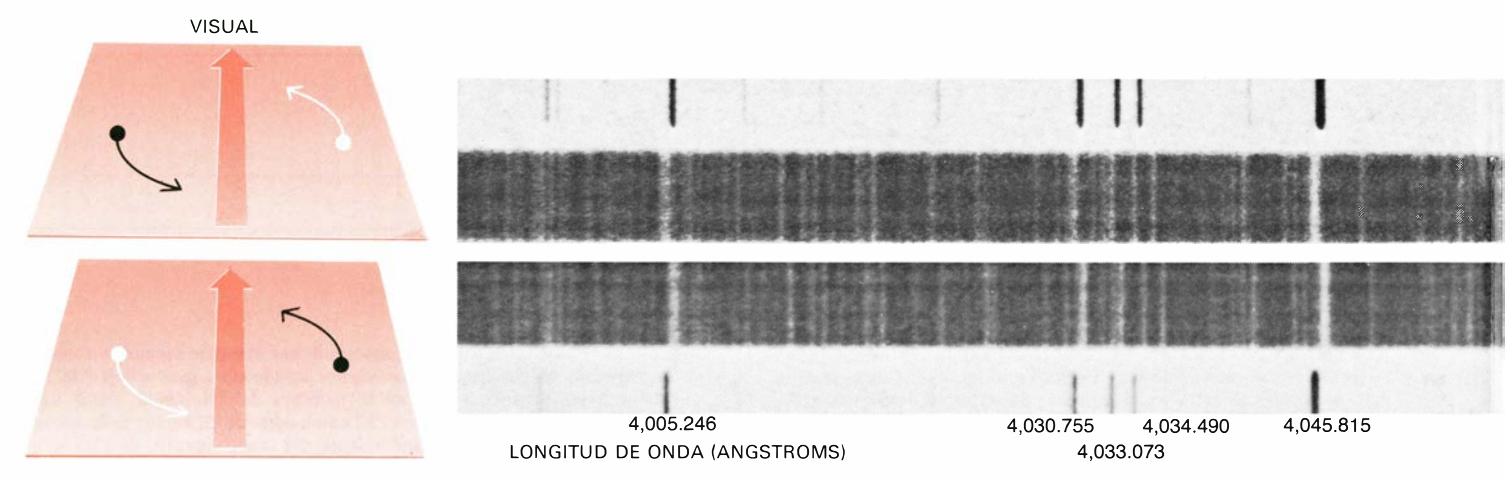
BINARIA ESPECTROSCOPICA de rayas dobles. Los espectrogramas son franjas horizontales de color *(bandas oscuras horizontales sobre negativos en blanco y negro)* **con huecos de absorción** *(tenues líneas verticales),* **atribuibles a la presencia de ciertos átomos en las atmósferas de las estrellas. En este caso, las rayas oscuras (positivo) de referencia para las rayas de absorción estelares fueron producidas por una lámpara de arco y corresponden a varias rayas de emisión asociadas a los átomos de hierro ionizados una sola vez (Fe I) y a los átomos de manganeso**

observó las 123 estrellas de tipo solar visibles, sin ayuda de instrumentos, desde el hemisferio norte. Cada una de esas estrellas está dentro de los 85 años-luz de distancia al sol, y parecía probable que, entre ellas, prácticamente todas las dobles visuales que pudieran detectarse con un equipo corriente hubieran sido ya registradas. Para muchas de esas estrellas, sin embargo, no había pruebas de corrimientos Doppler variables. En consecuencia, obtuvimos unos 20 espectrogramas de cada estrella con el gran espectrógrafo instalado en el foco *coudé* del telescopio reflector de 2,1 metros (84 pulgadas) de Kitt Peak. Los espectrogramas se hicieron en épocas distribuidas al azar, puesto que no sabíamos de antemano qué estrellas en particular eran dobles y, de serlo, si sus períodos orbitales se medirían por días, meses o años. Midiendo los corrimientos Doppler en tiempos tomados al azar pudimos buscar y, normalmente, encontrar el período que hace que los desplazamientos Doppler se repitan cíclicamente. Dentro de nuestra muestra de 123 estrellas, descubrimos 25 nuevas dobles espectroscópicas y mejoramos los elementos orbitales de la mayoría de las 21 dobles espectroscópicas anteriormente conocidas. Había además, en la muestra, 23 dobles visuales con elementos orbitales conocidos y un total de 25 parejas con movimiento propio común. Algunos pares se detectaron a través de más de un método, con lo que el total neto fue de 88 parejas.

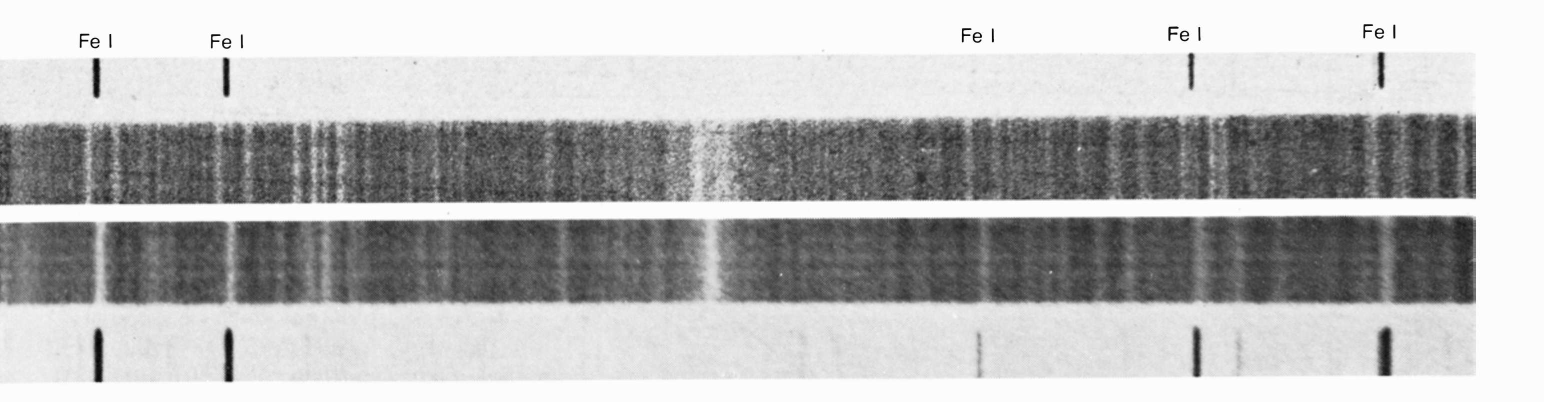
Algunos de los sistemas observados eran triples, e incluso cuádruples. Un ejemplo típico lo constituiría una pareja visual, una de cuyas componentes era una doble espectroscópica, lo que hace un total de tres estrellas aunque la acompañante espectroscópica no se vea en realidad. De entre las 123 estrellas primarias, para 57 se encontró que tenían un acompañante; para 11, dos acompañantes, y tres tenían tres acompañantes. En tantos por ciento, las frecuencias observadas de estrellas aisladas dobles, triples y cuádruples eran, respectivamente, el 42, el 46, el 9 y el 2 por ciento. Consiguientemente, más de la mitad (el 57 por ciento) de las estrellas tenían, por lo menos, un acompañante estelar.

El inventario anterior es sólo una parte del cuadro, puesto que no estamos capacitados todavía para detectar todos los acompañantes que probablemente estén presentes. Por ejemplo, si una pareja de estrellas estuviera tan junta que las componentes no pudiesen detectarse como una pareja visual y sucediera que el plano de las órbitas fuese exactamente perpendicular a la visual, las estrellas, al moverse en sus órbitas, no se acercarían ni se alejarían apreciablemente. Los corrimientos Doppler serían, por tanto, virtualmente constantes, con independencia de la velocidad con que las estrellas describieran sus órbitas. No podríamos detectar semejantes parejas "boca arriba".

Sin embargo, puede adelantarse cuántas de esas parejas "boca arriba" se dejaron. Hay pruebas sólidas de que los planos orbitales de las dobles espectroscópicas están orientados al azar. La mejor prueba viene de determinar en varias partes del cielo, el porcentaje de estrellas que son dobles eclipsantes. Su-Shu Huang y Clarence Wade, Jr., de la Northwestern University, encontraron que la fracción



BINARIA ESPECTROSCOPICA de rayas simples, llamada ι Pegasi; puede detectarse en este par de espectrogramas pese a que el espectro de la estrella secundaria no aparezca en absoluto. En el espectrograma superior, las rayas de absorción están desplazadas hacia la izquierda de las de referencia; ello indica que esa noche la estrella primaria, visible, se estaba acercando. En el espectrograma inferior, las rayas estelares



ionizados una vez (Mn I). En el espectrograma superior, las rayas de absorción son dobles porque una de las componentes del sistema se está acercando mientras la otra se va alejando. (Según el efecto Doppler, las rayas espectrales de un objeto que se acerca están desplazadas hacia el extremo violeta del espectro, mientras que las rayas de otro que se va alejando están corridas hacia el extremo rojo.) En el espectrograma inferior, ambas estrellas se están moviendo en un plano perpendicular a la visual, por lo que ambos conjuntos de rayas coinciden.

de dobles eclipsantes es la misma en todas direcciones, o sea, que los planos orbitales están orientados al azar. Se puede, pues, calcular estadísticamente cuántas dobles dejamos de detectar porque sus planos orbitales sean perpendiculares a las visuales. Probablemente, se nos escapó una sola pareja.

Lo que es más importante: hay muchas dobles que no pueden detectarse porque la masa de la compañera es demasiado pequeña para producir corrimientos Doppler de variabilidad apreciable en la estrella primaria. Las mismas observaciones muestran ya que si la velocidad orbital de la primaria es inferior a unos dos kilómetros por segundo, resulta muy difícil distinguir una variación real de un efecto producido por la imprecisión de las mediciones. Las velocidades orbitales típicas para las dobles espectroscópicas conocidas anteriormente se escalonan entre los 10 y los 75 kilómetros por segundo; para las dobles descubiertas ahora, las velocidades orbitales se sitúan entre los dos y los diez kilómetros por segundo.

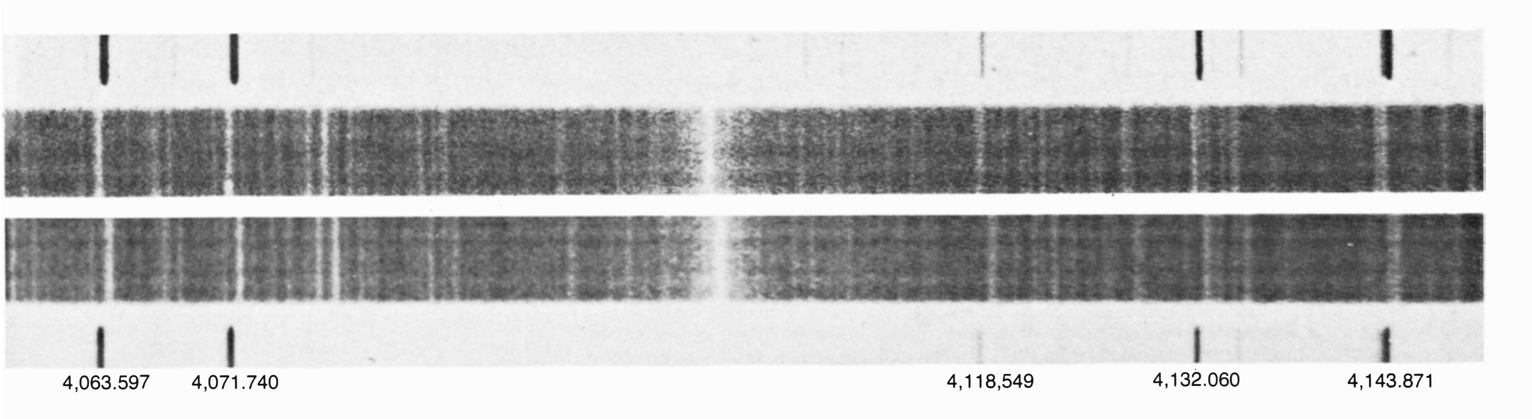
De nuevo, resultó que, tras considerar las dobles que detectamos, pudimos estimar cuántas se nos escaparon en razón de la masa pequeña de las acompañantes. Para sistemas en los cuales la acompañante tiene la mitad de la masa de la estrella primaria, se nos escapó, probablemente, menos de una pareja espectroscópica, frente a los 12 sistemas registrados; para sistemas con acompañantes que tienen un octavo de la masa de la estrella primaria, nos dejamos seguramente dos parejas espectroscópicas frente a las seis que descubrimos.

Existen, por supuesto, otras parejas de estrellas que escaparon a la detección; por ejemplo, parejas muy separadas fácilmente resolubles con un telescopio, en la que el acompañante despida tan poca luz que se pierda entre la multitud de estrellas de fondo, débiles, que rodean a cualquier estrella brillante. En vez de entrar en más detalles sobre que tipos de parejas que faltan pueden predecirse y tenerse o no en cuenta, en nuestro propósito por contar todas las parejas en una muestra de estrellas, describiré los resultados conseguidos.

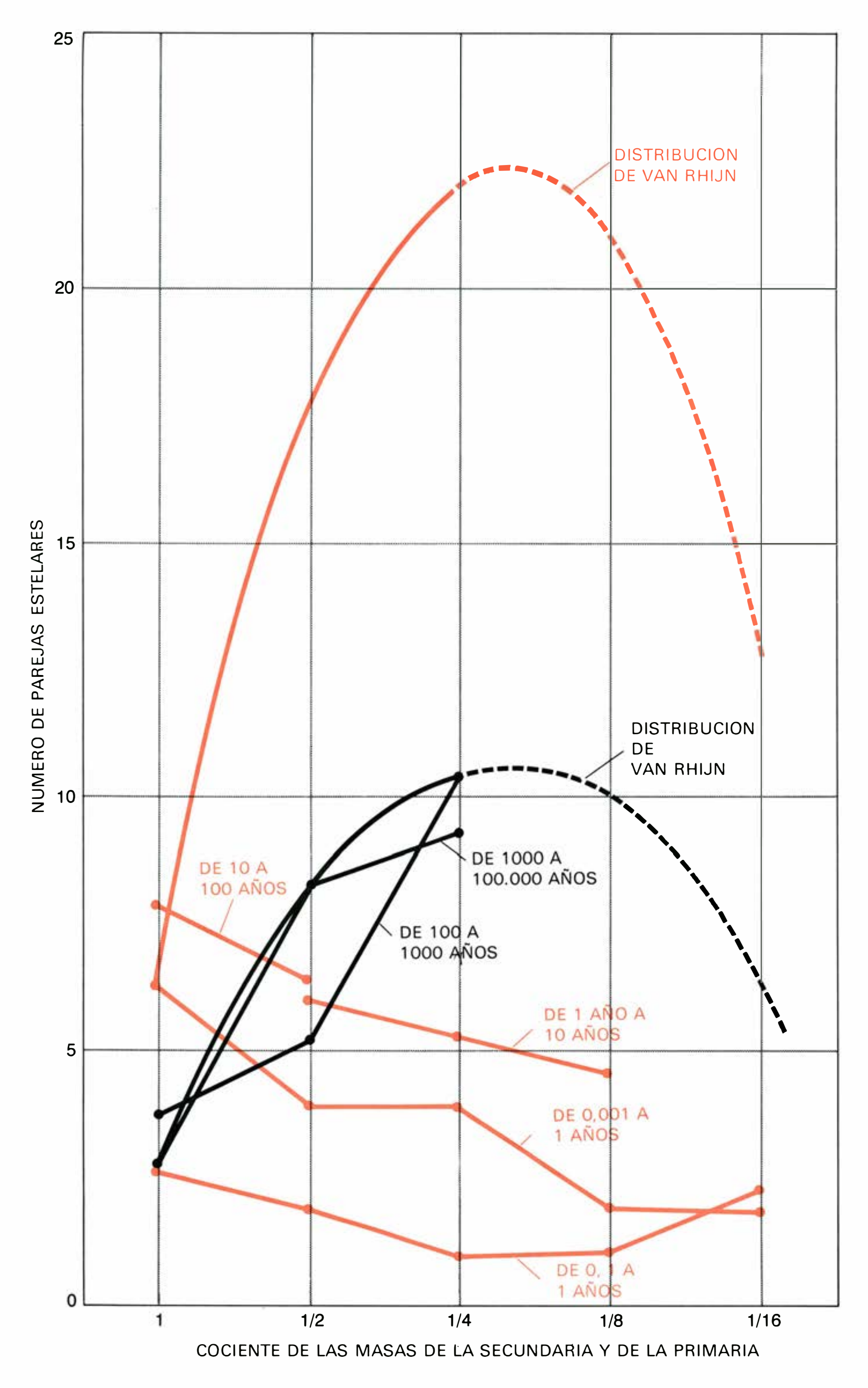
En nuestra muestra de 123 estrellas primarias, contamos el número de acompañantes, ya sea observados, ya sea deducidos de los cálculos de defectividad, con masas distintas. Combinamos las parejas en grupos: primero, las de masas más o menos iguales; luego, las parejas con estrellas secundarias de aproximadamente la mitad de la masa de las primarias, etcétera. Separamos también las parejas según sus períodos orbitales. Los resultados se dispusieron en una gráfica *(véase la ilustración de la página 68)*.

En la gráfica, los resultados para parejas con períodos orbitales de menos de 100 años aparecen en color; los resultados para parejas con períodos más largos, en negro. Faltan algunos puntos, debido a que, en determinadas circunstancias, no pudimos calcular con exactitud el número de parejas no detectadas.

A partir de la gráfica, y para las parejas con un período orbital de menos de 100 años, yendo de las parejas con masas casi iguales a las parejas en que la estrella secundaria tiene una masa pequeña, se encuentran obviamente cada vez menos sistemas. Los resultados para las



están corridas hacia la derecha; significa que la estrella primaria se estaba alejando esa noche. Los desplazamientos miden unos 0,04 milímetros en los espectrogramas originales y corresponden a un movimiento orbital de unos 40 km/seg. Los cuatro espectrogramas, hechos con el espectrógrafo *coudé* del telescopio de Kitt Peak, muestran parte de la región violeta del espectro; las longitudes de onda más largas están a la derecha.



RESULTADOS DEL ESTUDIO ESPECTROGRAFICO del autor sobre la frecuencia de los sistemas estelares múltiples entre las 123 estrellas de tipo solar observables a simple vista desde el hemisferio norte. Primero, los acompañantes estelares observados (y deducidos) se clasificaron de acuerdo con las masas. Las parejas con masas aproximadamente iguales se combinaron en un grupo; las parejas con estrellas secundarias de la mitad de la masa de las primarias, aproximadamente, se combinaron en otro grupo, y así sucesivamente. Las parejas se dividieron entonces según sus períodos orbitales. Las parejas con períodos inferiores a los 100 años aparecen en color; las parejas con períodos más largos, en negro. Para las parejas con períodos orbitales de menos de 100 años, si se procede desde las parejas con masas más o menos iguales a las parejas con secundarias de masas pequeñas, se registran cada vez menos sistemas, mientras que, para las parejas con un período orbital de más de 100 años, yendo hacia las masas inferiores se encuentran más parejas, al menos hasta las secundarias con un cuarto de la masa de las primarias. Para las parejas en negro, las estrellas primarias y las secundarias parecen seguir (dentro de los límites de precisión de las mediciones) una curva regular llamada distribución de van Rhijn; no lo hacen las parejas en color, lo cual indica que hay dos tipos de sistemas estelares dobles: estrellas que se formaron juntas por parejas y estrellas que se formaron separadamente. La extrapolación de los datos aquí presentados, para ambos tipos de sistemas de estrellas dobles, sugiere que las 123 estrellas primarias de la muestra tienen todas acompañantes: el 67 por ciento de las secundarias son estrellas normales, un 15 por ciento son enanas negras y un 20 por ciento de dichas secundarias son planetas.

parejas con un período de más de 100 años son opuestos; cuando se va hacia las masas más pequeñas se encuentran más parejas, al menos hasta llegar a las secundarias cuya masa es un cuarto de la masa de la primaria.

Las propias estrellas primarias siguen una curva regular llamada distribución de van Rhijn, en memoria del astrónomo holandés Pieter J. van Rhijn, el cual hizo recuentos de estrellas en zonas de muestra del cielo, cuyas distancias individuales eran conocidas. Encontró, por ejemplo, que si en un volumen dado del espacio próximo al sol había seis estrellas de una masa solar, aproximadamente, habrá 18 con la mitad de esa masa, 22 con un cuarto de masa, y luego, tal vez, números decrecientes con masas aún menores. Los datos de van Rhijn probablemente no eran muy buenos para estrellas cuya masa sea inferior a un cuarto de masa solar. De ahí que, a partir de ese punto, las curvas resulten discontinuas. Para las parejas con un período superior a 100 años, la curva está ajustada arbitrariamente para que empiece con 2,7 estrellas de alrededor de una masa solar. Aparentemente, en esas parejas, tanto las primarias como las secundarias siguen la distribución de van Rhijn, mientras que las parejas con períodos inferiores a 100 años no lo hacen.

Esos datos sugieren que hay dos tipos de sistemas de estrellas dobles. Antes de examinar el significado de este descubrimiento, será útil pasar revista a parte de lo que se sabe sobre la formación de las estrellas.

Las estrellas muy jóvenes pueden conocerse de varias maneras. Una de las más fáciles estriba en identificar una estrella que esté quemando su combustible nuclear con tal rapidez que su tiempo de vida como estrella enana no llegue al millón de años. Tales estrellas se encuentran casi siempre en nubes de gas, como la Gran Nebulosa de Orión. En raras ocasiones, se las encuentra también viajando rápidamente a través del espacio, alejándose de sus puntos de formación. Se las suele designar como "estrellas en despegue". A partir de la presencia observada de estrellas jóvenes en nubes de polvo y gas y de los cálculos teóricos, se deduce claramente que las estrellas se originan en nubes densas. El mecanismo de formación sólo está explicado en parte. Es necesario que porciones de las nubes de gas alcancen una densidad crítica tal que la fuerza atractiva de la autogravitación supere a las fuerzas dispersoras de la presión y la turbulencia del gas, así como a las fuerzas ejercidas por campos magnéticos, que inciden en los movimientos de las partículas cargadas. Par-

tes de una nube pueden alcanzar una tal densidad, debido a colisiones accidentales, o a una onda de choque procedente de la explosión de una supernova cercana. En cualquier caso, cuando las condiciones son adecuadas, se genera todo un cúmulo de estrellas de una sola vez.

Las partes densas de las nubes de gas poseen mucho movimiento interno, o turbulencia, que no se elimina fácilmente. Cuando una protoestrella, o parte de una nube que se condensa, se contrae, suele girar con mayor celeridad, llegando con el tiempo al punto en que se hace inestable. La protoestrella puede evitar, en parte, esa suerte eyectando las partículas que tengan el mayor momento angular, pero no puede llegar a eliminar totalmente la rotación rápida. Si una nube con una densidad media de 100 átomos por centímetro cúbico produce estrellas con una densidad media de 10^{23} átomos por centímetro cúbico, el movimiento interno medio del gas de la nube, que es de alrededor de un kilómetro por segundo, aumentaría hasta los 10 millones de kilómetros por segundo. Semejante velocidad es 20.000 veces mayor que la más alta velocidad de rotación posible para una estrella estable. De nuevo, el separar y expulsar los átomos que tengan mayor momento angular no resolvería enteramente el problema. ¿Cómo puede formarse, si es que puede, una estrella aislada estable?

Cuando una protoestrella se contrae y gira con mayor rapidez, su forma se aparta de la configuración más o menos esférica. Primero, se convierte en un esferoide alargado o aplanado; luego, a velocidades aún mayores, se transforma en un elipsoide con tres radios desiguales. Por último, el objeto se divide en dos elipsoides separados, proceso de bifurcación que conduce a una estrella doble con una separación pequeña. En el proceso, aproximadamente el 99 por ciento del movimiento angular va a parar al movimiento orbital; sólo alrededor del 1 por ciento va a la rotación de ambas estrellas. De donde resulta que, dividiéndose en dos y perdiendo algunos de sus átomos más rápidos, una protoestrella en rotación rápida puede desarrollarse como una pareja de estrellas estables.

El mecanismo arriba descrito parece ser el que se aplica a las estrellas de la primera categoría, que son aquellas parejas con períodos inferiores a los 100 años. La frecuencia decreciente de las estrellas dobles que tienen una secundaria con una masa menor, indica que, en el proceso de bifurcación, la tendencia dominante es que la protoestrella se divida en una pareja de estrellas con masas iguales. Las parejas con masas muy diferentes abundan menos.

¿Cuántas parejas de ésas, de dobles de bifurcación, existen en nuestra muestra de 123 primarias? En la gráfica, todas las líneas quebradas de color tienen, aproximadamente, la misma pendiente, dentro de la precisión de nuestras mediciones para una muestra tan reducida de estrellas. (Nos hubiera gustado haber empezado con una muestra mayor, pero las observaciones de las 123 estrellas llevaron 60 noches de tiempo de observación con un gran telescopio.) Si se extrapolan esas líneas y se cuentan todas las secundarias cuyas masas van desde la misma masa que la primaria hasta un dieciseisavo de la masa de la primaria, se observa que hay 83 secundarias próximas entre las 123 primarias; es decir, que dos tercios de las primarias tienen secundarias cercanas.

Una masa igual a un dieciseisavo de la masa de sol es la cantidad mínima de materia que puede originar una estrella que genere luz a partir de reacciones nucleares en su interior. Este importante hecho fue descubierto en 1963 por Shiv S. Kumar, de la Universidad de Virginia. Si una esfera gaseosa tiene una masa inferior, el peso de esa masa, al apretar el núcleo central, no llegará a producir una temperatura lo bastante alta como para que tengan lugar reacciones nucleares. En el sol, el peso del material hace que la temperatura del gas sea de unos 14 millones de grados Kelvin en el centro, temperatura a la cual los núcleos atómicos (mayormente protones) tienen una velocidad media de 100 kilómetros por segundo. Los protones más rápidos tienen una velocidad suficiente como para que, cuando dos de ellos siguen una trayectoria de colisión, choquen efectivamente pese a la repulsión debida a sus cargas eléctricas iguales. Tras la colisión pueden formar un núcleo de deuterio y, después de colisiones adicionales, un núcleo de helio. Un núcleo de helio pesa algo menos que los cuatro núcleos de hidrógeno que lo han formado; el material perdido aparece como energía, según la relación $E = mc^2$ (la energía generada es igual a la masa perdida multiplicada por el cuadrado de la velocidad de la luz). En cambio, en una esfera gaseosa con una masa igual o menor que un dieciseisavo de masa solar la temperatura central no llega al millón de grados Kelvin, y los protones se desplazan con una velocidad media de sólo 25 kilómetros por segundo. Prácticamente ninguno de esos protones lleva una velocidad lo bastante grande como para que dos protones en trayectoria de colisión venzan su repulsión eléctrica. Los protones se acercarán el uno al otro, frenarán y volverán atrás, sin tener nunca la posibilidad de chocar y formar un núcleo más pesado.

¿Qué les sucede a esos objetos, que inician su curso para convertirse en estrellas, pero nunca lo hacen porque contienen poco material? Kumar trazó su historia y encontró que los que tienen masas comprendidas entre un dieciseisavo y una centésima de masa solar brillarán durante unos mil millones de años, mientras se contraen, denido a que convierten energía gravitatoria en energía electromagnética. Pasado ese tiempo, seguirán existiendo indefinidamente, como enanas negras no luminosas (que no hay que confundir con los agujeros negros).

Una esfera de gas con una masa inferior a una centésima de masa solar difícilmente llegará a brillar en absoluto (y, en tal caso, sólo en longitudes de onda infrarrojas). Semejante objeto –una esfera de gas cuya masa resulta demasiado pequeña como para convertirse en una estrella, ni siquiera en una enana negra– se llamará un planeta. Si un cuerpo de esta especie tiene una masa muy pequeña, poseerá una superficie sólida.

Es improbable que las líneas quebradas de colores de la gráfica caigan a cero para cuerpos secundarios cuyas masas no alcancen un dieciseisavo de la masa solar. Las líneas probablemente continúan, indicando que algunas secundadarias serán enanas negras o planetas.

Al principio pensamos que la pendiente media de las líneas de colores de la gráfica era tal que el número de sistemas dobles resultaba proporcional a la raíz cúbica de la masa de la secundaria. Por ejemplo, a un octavo de masa solar habría la mitad de sistemas dobles que a una masa solar: $(1/8)^{1/3} = 1/2$. David R. Branch, de la Universidad de Oklahoma, señaló dos efectos que habían sido pasados por alto; de ellos resultaba que la relación con la raíz cúbica podría ser otra en realidad, posiblemente una relación con la raíz cuadrada. Dicho de otro modo, cualquier extrapolación a partir de los presentes datos peca de arriesgada. Físicamente, ello significa que no hemos observado dobles con enanas negras o planetas como acompañantes secundarios, y que nuestros datos para las secundarias de masas mayores no son lo bastante seguros como para extrapolar con confianza.

Sin embargo, si se extrapola en base a la relación con la raíz cúbica, los resultados tienen sentido. Si prolongamos las curvas de colores, primero hasta una centésima de masa solar y luego hasta

masa cero, se observará que hay unas 20 secundarias que son enanas negras y otras 25 que son planetas. La suma de éstas, más las 83 secundarias estelares, resultan casi exactamente 123 ¡lo cual significa que todas las primarias tienen secundarias! En un 67 por ciento de los casos, las secundarias son estrellas normales, en un 15 por ciento son enanas negras y en un 20 por ciento son planetas. La razón de que este resultado tenga sentido, pese a la endeble extrapolación en que se basa, es que significa que todas las primarias son dobles de bifurcación; el problema de formar estrellas estables pese al exceso de movimiento angular se resuelve, en líneas generales, yendo a parar la mayor parte de éste al movimiento orbital.

El sol está dentro del 20 por ciento de estrellas de la muestra que tienen planetas por acompañantes. El hecho de que, en el sistema solar, la secundaria se haya desarrollado formando un conjunto de planetas, en vez de un planeta gigante, no altera de manera significativa el modelo. Concluimos, pues, como lo han hecho otros sobre una base más intuitiva, que en la búsqueda de planetas habría que concentrarse en las estrellas de las que no se sabe que tengan otras por acompañantes.

¿Qué puede decirse de las parejas muy separadas, es decir, de las parejas representadas por las curvas negras en la gráfica de nuestros resultados? Está claro que una pareja con un período orbital de más de 100 años no es el resultado de una bifurcación de una protoestrella en rotación rápida. Semejante pareja separada es, probablemente, el resultado de dos protoestrellas que se contrajeron por separado, y que se hallan ligadas una a la otra por la gravedad. Puesto que representan dos partes de una nube de gas, pueden empezar con cualquier cantidad de éste. La distribución de frecuencias de sus masas se aproxima a la distribución de van Rhijn, tanto para las primarias como para sus lejanas acompañantes. La manera según la cual una nube de gas turbulento, sometida a las fuerzas de autogravitación, presión interna y ondas de choque interestelares, se fragmenta en estrellas de distintas masas, da lugar a la distribución de van Rhijn, aunque en la actualidad los astrónomos sean incapaces de predecir esa distribución a partir de la teoría. El hecho a resaltar aquí es que la tendencia a formar masas primarias y secundarias aproximadamente iguales, que surge de la bifurcación, no se muestra clara entre las dobles separadas.

Si se extrapolan los datos representados en negro en la gráfica, utilizando la distribución de van Rhijn, se obtiene que unos tres cuartos (72 por ciento) de las primarias tienen acompañantes estelares lejanos (primarias con masas pequeñas, del orden de un dieciseisavo de la masa del sol). Si se conociera la distribución de van Rhijn más allá de un dieciseisavo de masa solar, la curva podría no caer a cero, sino abarcar la formación de enanas negras y de planetas. Podría resultar que la mayoría de las primarias tuvieran acompañantes lejanos y cercanos. ¿No sería verdad que las protoestrellas alejadas tendrían el mismo problema de poseer un momento angular excesivo como para formar una estrella estable, problema que sólo puede resolverse por bifurcación? Parece que sí. Yo adelantaría que los acompañantes lejanos son, en realidad, parejas juntas de bifurcación, igual que las primarias. En otras palabras, la situación más probable es una pareja junta de bifurcación, gravitatoriamente ligada a otra pareja de bifurcación, distante, es decir, un sistema cuádruple. Dentro de cada pareja, el acompañante de menor masa puede ser o bien una estrella (en las dos terceras partes de los casos, aproximadamente), una enana negra (en un sexto de los casos), o un planeta (en un sexto de los casos, más o menos).

El análisis que precede sugiere un cuadro atractivo, en el que la mayoría de las estrellas se formarían como miembros de sistemas múltiples. Tal resultado no es completamente inesperado; desde hace ya cierto tiempo, se supone que las estrellas aisladas del cielo se escaparon de cúmulos que se fueron desintegrando gradualmente. Esta sospecha proviene de la dificultad teórica que implica tratar de explicar cómo una nube aislada de una masa solar, por ejemplo, se contrae para formar una estrella. Se sabe asimismo que la mayoría de los cúmulos abiertos son gravitatoriamente inestables e irán perdiendo una a una sus estrellas, a lo largo de los millones de años de su existencia.

El cuadro no se sale del marco de la sugerencia, pues los datos presentados no son ni lo bastante numerosos ni lo bastante depurados como para estar seguros de una extrapolación que llegue hasta las enanas negras y los planetas. La conclusión ha de esperar a que el desarrollo de telescopios espaciales o un mejor equipamiento de los terrestres proporcionen información mucho más firme sobre la frecuencia de los planetas. Mientras tanto, nada nos impide fantasear sobre las condiciones en esos planetas y especular sobre si las señales de radio de la Tierra son recogidas en el otro extremo por un oyente, o incluso incitan a un locutor.

Evolución cromosómica de los primates

El estudio de los cromosomas de los primates revela los posibles mecanismos que han dado origen a la especie humana y abre un amplio campo a la investigación de las enfermedades hereditarias

José Egozcue

Los primates, a cuyo orden zoológico pertenece el hombre, han despertado siempre la curiosidad y el interés científico, tal vez debido al estrecho parentesco que a ellos nos une en la escala animal y a la similitud observada en muchas reacciones y formas de comportamiento. Así, por ejemplo, en las antiguas ilustraciones de antropoides, como las que aparecen en las obras de Buffon, el artista "humaniza" inconscientemente los rasgos de los chimpancés y de los orangutanes, al tiempo que "maligniza" a los gorilas, a causa del imponente y amenazador aspecto que estos tímidos animales presentan.

Con la entrada en el campo de la ciencia de las ideas darwinistas sobre la evolución de las especies aparece una obsesión tenaz y conceptualmente errónea por identificar lo que se denominaba el "eslabón perdido". El afán de los paleontólogos era hallar la secuencia exacta seguida por la evolución de los primates, que esperaban encontrar tras cada descubrimiento realizado en los yacimientos africanos y asiáticos.

El desarrollo de la ciencia permitió avanzar más. De las clasificaciones fundadas en las características morfológicas y morfométricas se pasó a un plano genético. No bastaba ya con las divergencias fenotípicas más aparentes, sino que era necesario incorporarles fenotipos subcelulares, como los cromosómicos, e incluso estudios genotípicos.

La genética, pues, se ha situado en un primer plano en el estudio de la evolución de los primates hasta llegar al hombre, sin prescindir, por ello, de otras disciplinas con las que se complementa para describir qué es el hombre y de dónde procede. La ciencia de la genética ha progresado muchísimo en nuestro siglo, desde el redescubrimiento de las leyes de Mendel. Y se ha completado la genética "mendeliana" o de la herencia de los caracteres con la genética "molecular", tras el descubrimiento del código genético y la identificación de los ácidos nucleicos fundamentales.

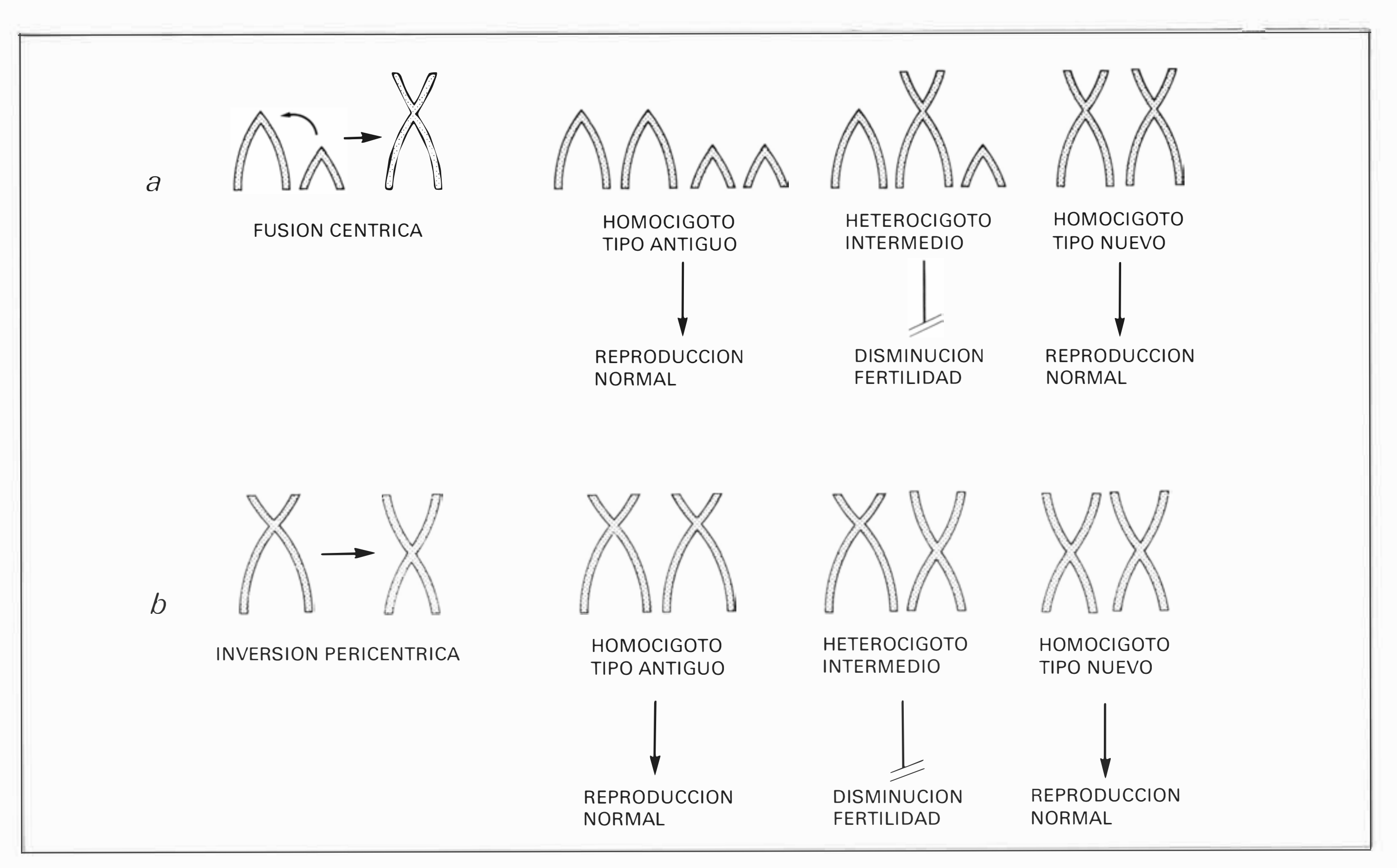
Tanto la genética mendeliana como la genética molecular estudian la herencia de los caracteres de los progenitores por parte de su descendencia y la variabilidad observada en esta transmisión. En los organismos superiores, esta variabilidad se debe en gran parte a la mezcla de los caracteres de los abuelos que se produce durante la formación de los gametos (ovocitos y espermatozoides) del padre y de la madre mediante intercambio entre sus cromosomas. Por otra, a consecuencia de la aparición de mutaciones, es decir, cambios en la secuencia de nucleótidos que forman el material hereditario o ADN.

El material hereditario se halla situado en los cromosomas. Los cromosomas son cuerpos filamentosos constituidos principalmente por ADN (ácido desoxirribonucleico) y proteínas que se encuentran en número variable en el núcleo de las células eucarióticas. Los cromosomas se presentan a pares, normalmente varios pares diferentes por núcleo. Los dos miembros de cada par tienen idéntico aspecto, y se dice que son homólogos entre sí. Los cromosomas homólogos se asocian o aparean de una manera característica durante el proceso de la meiosis.

Los cromosomas presentan una estructura bien definida. Su longitud sirve para numerarlos: cromosoma 1, cromosoma 2, etcétera. Los cromosomas sexuales no se numeran, sino que se simbolizan así: X e Y; en el sexo masculino (cromosomas no homólogos) el par es XY, siendo el primero de procedencia materna y el segundo de procedencia paterna; los individuos hembra poseen dos cromosomas X (homólogos). La caracterización morfológica de los cromosomas se basa en su longitud y en la posición de una constricción bien definida llamada centrómero.

El número de cromosomas que se hallan en el núcleo es constante para cada especie en un momento dado de su historia evolutiva. Y esto es una razón para distinguir entre las distintas especies de primates afines al hombre. El número total de cromosomas puede encontrarse en un estado haploide, diploide o poliploide. Un organismo, una célula o un núcleo se llama haploide cuando posee un conjunto de cromosomas no apareados, que simbolizaremos por n. Será diploide cuando posea el juego cromosómico apareado, cuyo simbolismo es $2n$, si el conjunto haploide cromosómico fundamental era n. Nos hallaremos ante una situación o un estado de poliploidía cuando siendo el número fundamental (o haploide) de cromosomas n, en un organismo determinado se multiplica dicho número por 3, 4, 5...; lo cual se simboliza, respectivamente por $3n$, $4n$, $5n$...

El gen, que constituye la unidad de la herencia, es el elemento fundamental de los cromosomas; cada gen forma, pues, una parte discreta del cromosoma. Los genes, según esta idea básica, pueden seriarse y constituir lo que se llama un mapa cromosómico, mapa que nos dará la imagen visual de cuáles sean las distintas semejanzas o diferencias entre dos primates, a un nivel profundo, que explican las otras diferencias aparentes a simple vista.



MECANISMOS CROMOSOMICOS de especiación. Tanto la fusión céntrica *(a)*, como la inversión pericéntrica *(b)*, originan problemas de segregación meiótica en el heterocigoto. Hacen que su fertilidad descienda y favorecen así los cruzamientos entre homocigotos de tipo "antiguo" y de tipo "nuevo", dando lugar a la formación de nuevas especies. Los miembros de la especie forman un grupo aislado en cuanto a la reproducción.

Los primeros estudios cromosómicos realizados en primates datan de 1922. De hecho, mucho antes de que se determinara con exactitud el número cromosómico de la especie humana se conoció el número cromosómico de ciertos monos sudamericanos.

Sin embargo, la verdadera explosión en este campo de la ciencia se produce coincidiendo con la creación, en diversos países del mundo, de centros de primatología en los que, por vez primera, se empiezan a utilizar primates de forma sistemática para la investigación biomédica. Profundizando en el estudio de los cromosomas de numerosas especies de primates, se obtienen datos más exactos sobre sus relaciones filogenéticas o evolutivas.

De esta forma se consigue establecer con mayor precisión el grado de relación existente entre distintas especies, y se demuestra también que la clasificación actual, basada principalmente en criterios anatómicos, adolece de importantes defectos. Así, por ejemplo, ciertos grupos de primates clasificados como *géneros* distintos presentan cromosomas idénticos y producen "híbridos" fértiles, por lo que, desde el punto de vista genético, no cumplen siquiera con el requisito necesario para ser clasificados como *especies* distintas. Un individuo pertenece a una especie si, en primer lugar posee los cromosomas que caracterizan a esa especie, pues, según se vio, los cromosomas difieren de una especie a otra. Es decir, los miembros de una especie forman un grupo aislado desde el punto de vista de la reproducción; sus genes no se combinan con los genes de otro grupo, pero sí con los de otros individuos que pertenezcan a esa especie, y originan, en su combinación, una descendencia fértil.

Por otra parte, en una misma especie de primates, según la clasificación tradicional, aparecen diferencias cromosómicas suficientes como para establecer una barrera de fertilidad entre diversos grupos que, al no poder reproducirse entre sí, según el criterio genético, deberían clasificarse como especies distintas.

Al mismo tiempo, y gracias a la contribución de diversos campos de la investigación biomédica, se pone de manifiesto que muchas de las enfermedades de tipo hereditario que afectan a la especie humana aparecen también, con características idénticas, en determinadas especies de primates. Como ejemplos más patentes podemos citar el síndrome de Down (trisomía del cromosoma número 21, es decir, en vez de dos cromosomas homólogos, el paciente ha nacido con tres cromosomas), que está perfectamente estudiado en el chimpancé *(Pan troglodytes)*, la alcaptonuria (enfermedad hereditaria debida a la ausencia de un gen que regula la oxidación del ácido homogentísico) que se ha podido describir en el orangután *(Pongo pygmaeus)* y otras muchas.

También se demuestra, en estas mismas fechas, que ciertos primates responden frente a determinados teratógenos de forma idéntica a como lo hace la especie humana, como ocurre con los babuinos frente a la talidomida o con los macacos rhesus frente a la d-dietilamida del ácido lisérgico (LSD).

Estas observaciones, que abren un campo muy prometedor al estudio de las enfermedades de origen genético y de las malformaciones congénitas, coinciden con la puesta a punto de las técnicas de hibridación celular. Mediante estas técnicas, al mezclar en una sola célula los cromosomas procedentes de dos especies distintas, se hace posible determinar con gran exactitud la localización de los genes en los cromosomas, y, en consecuencia, la obtención de mapas génicos detallados de la especie analizada.

Por consiguiente, la posibilidad de estudiar en un animal de experimentación una enfermedad que está determinada por un gen cuya localización en el cromosoma *(locus)* se conoce con exactitud

facilita extraordinariamente el estudio de su etiología y posible tratamiento.

Debido a ello, hoy se están dedicando grandes esfuerzos a la obtención de mapas génicos en aquellas especies que cumplen con las dos condiciones básicas siguientes: 1) poder ser utilizadas como modelos para la investigación de enfermedades hereditarias que afecten a la especie humana, y 2) reproducirse adecuadamente en cautividad, de forma que su empleo en programas de investigación no ponga en peligro la viabilidad de la especie. De esta forma, se pueden usar primates para la investigación y respetar, al mismo tiempo, las normas de protección, no sólo para las especies amenazadas de extinción, sino también para cualquier otro tipo de primate.

Los estudios realizados hasta el momento han permitido obtener importantes datos con respecto a los mecanismos cromosómicos de especiación.

Los mecanismos de especiación, que resultan muy eficaces en poblaciones aisladas, o en poblaciones periféricas semiaisladas, con un alto nivel de endogamia (reproducción limitada a un círculo restringido que conduce a la consanguinidad), dan lugar a la aparición de una barrera de fertilidad. Dicha barrera acaba por separar reproductivamente a la población original en dos grupos, que constituyen el punto de origen de nuevas especies.

Los mecanismos de especiación más eficaces, a nivel cromosómico, son, por este orden, la inversión pericéntrica y la fusión céntrica *(véase la ilustración de la página 73)*. La inversión es un tipo de mutación cromosómica en la cual un segmento de un cromosoma se rompe, gira 180° y vuelve a unirse. Por ejemplo, si un cromosoma consta de los siguientes genes ABCDEFGH, y se rompe un segmento que conste de los genes CDE, diremos que hay inversión cuando el cromosoma vuelve a restablecerse con el siguiente mapa génico: ABEDCFGH. La fusión céntrica es la soldadura de segmentos cromosómicos en forma de V en un cromosoma único, adoptando frecuentemente la forma de H.

La incidencia de estos cambios cromosómicos en la meiosis (forma de división celular que comporta el apareamiento exacto de los cromosomas homólogos y su separación posterior para producir células hijas haploides) depende de que se encuentren en homocigosis o en heterocigosis.

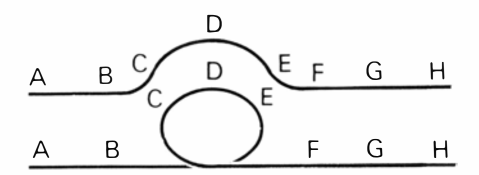
A nivel cromosómico decimos que un individuo es homocigoto cuando los dos cromosomas de un par presentan las mismas características, es decir, p. ej., no hay inversión en ninguno de los dos, o ambos la presentan. De esta forma, el apareamiento meiótico no plantea problemas:

$$\frac{\text{A B C D E F G H}}{\text{A B C D E F G H}} \quad \text{homocigoto normal}$$

o

$$\frac{\text{A B E D C F G H}}{\text{A B E D C F G H}} \quad \text{homocigoto para la inversión}$$

En cambio, el heterocigoto presenta un cromosoma de cada tipo, lo que dificulta el apareamiento meiótico que, en este caso sólo es posible si se forma un asa de inversión:



Las inversiones pericéntricas no afectan al número cromosómico básico de la especie, que no varía, pero sí pueden provocar cambios en la morfología del par cromosómico que presenta la inversión. En cambio, las fusiones céntricas dan lugar a que el número cromosómico inicial disminuya en dos unidades, pasando, p. ej., de $2n=48$ a $2n=46$.

En ambos casos, aparecen inicialmente individuos que son heterocigotos para el cambio, los cuales dan posteriormente origen a individuos homocigotos correspondientes a la nueva fórmula cromosómica.

Los heterocigotos presentan importantes dificultades reproductivas debido a que durante la meiosis el apareamiento de los cromosomas afectos es anormal, lo que da lugar a la formación de un elevado número de gametos desequilibrados y, en consecuencia, disminuye su fertilidad, ya que los productos de la concepción no son viables.

En cambio, los homocigotos correspondientes al tipo cromosómico antiguo y al tipo cromosómico nuevo se reproducen normalmente.

El cruzamiento de un homocigoto de tipo antiguo con un homocigoto de tipo nuevo da lugar al nacimiento de un individuo heterocigoto, que presentará dificultades reproductivas. En consecuencia, se ven favorecidos los cruzamientos entre individuos que presentan el mismo tipo cromosómico, lo que hace que, progresivamente, la forma cromosómica antigua y la forma nueva vayan distanciándose, hasta constituir nuevas especies.

Desde el punto de vista meiótico, el efecto que ejercen las inversiones pericéntricas es mucho más drástico que el que producen las fusiones céntricas, ya que, en general, el número de gametos anormales es mucho mayor en los heterocigotos para una inversión que en los heterocigotos para una fusión céntrica. Sin embargo, durante la evolución de los primates se han producido más fusiones que inversiones. Es posible que ello se deba a que las inversiones dan lugar a un número tan elevado de gametos anormales que su fijación en homocigosis resulta improbable.

Hasta hace pocos años, la identificación de estos mecanismos de especiación resultaba difícil y en gran parte intuitiva, ya que las técnicas existentes no permitían la identificación exacta de los pares cromosómicos. En consecuencia, era generalmente necesario intentar adivinar la forma en que se habían producido las diferencias cromosómicas existentes en las distintas especies.

En 1971, la aparición de técnicas de tinción específica, que permiten visualizar en los cromosomas bandas de intensidad de coloración variable, permitió llegar a una caracterización precisa de los cambios cromosómicos producidos en cada caso. Es interesante señalar, sin embargo, que las predicciones establecidas con anterioridad a esta fecha se han visto confirmadas en su casi totalidad después de la aplicación de las técnicas de bandas.

Los métodos más frecuentemente utilizados son los de bandas *Q* (tinción con un derivado de la quinacrina y observación al microscopio de fluorescencia), bandas *G* (tratamiento con enzimas proteolíticos o con soluciones salinas calientes y tinción con el colorante de Giemsa), bandas *R* ("reverse", o invertidas, obtenidas también mediante tratamiento con soluciones salinas) y bandas *C* (centroméricas, correspondientes a las regiones de ADN ribosómico). La introducción de estas técnicas ha permitido comprobar también que la organización del material genético es extraordinariamente estable, especialmente en ciertos cromosomas, lo que facilita la identificación de cromosomas con bandas idénticas en especies filogenéticamente muy lejanas, como pueden ser el prosimio *Nycticebus coucang* y *Homo sapiens*.

Por supuesto, es posible que los cromosomas que presentan un patrón de bandas idéntico contengan genes muy distintos, y que la disposición igual de sus bandas sea puramente casual. Sin embargo, los estudios de localización

génica por hibridación celular realizados hasta el momento indican que, en principio, a una identidad de bandas corresponde una identidad de función, es decir, que los cromosomas con bandas idénticas contendrían los mismos genes.

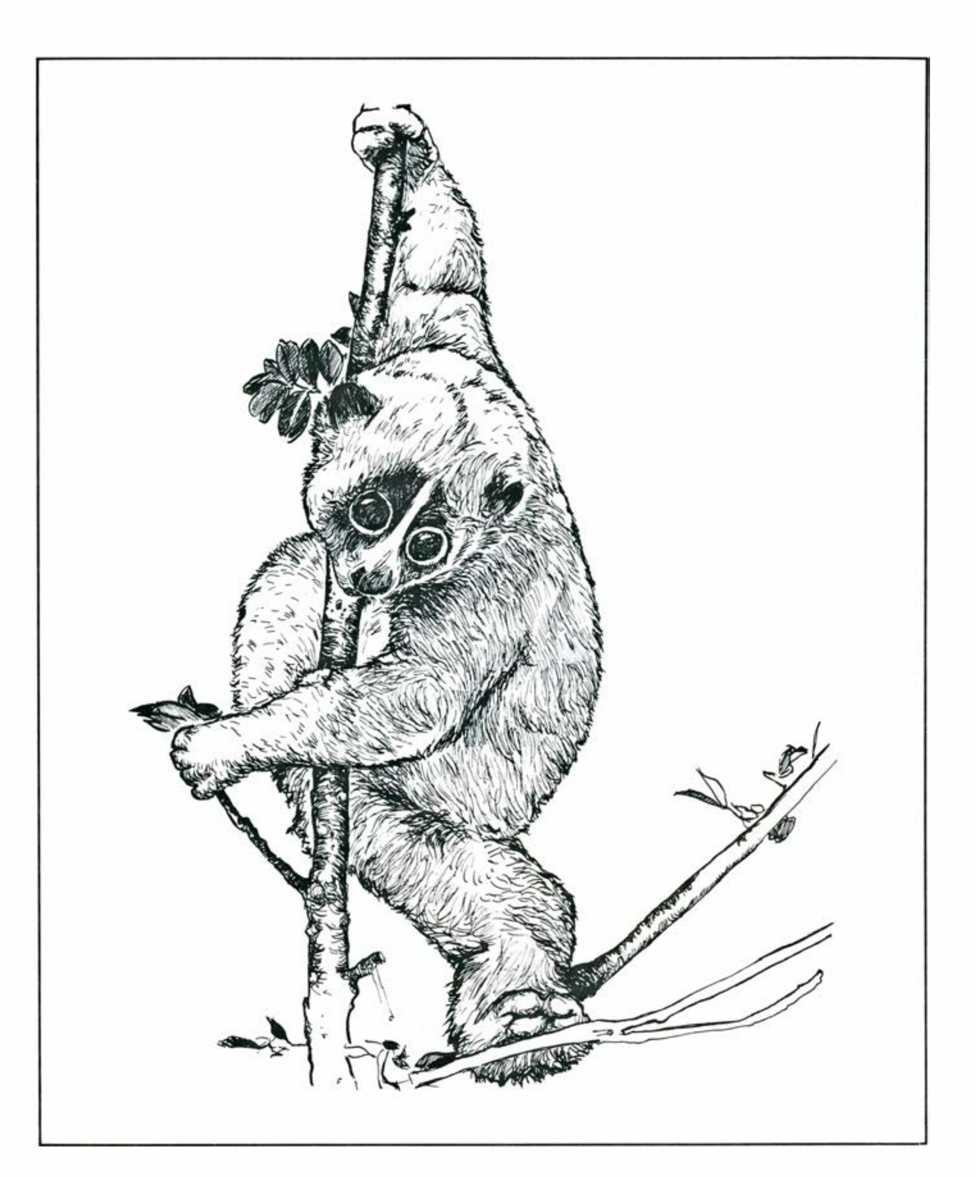
En conjunto, los mecanismos cromosómicos de especiación han tenido una incidencia muy variable sobre los distintos grupos de primates y también sobre los distintos pares cromosómicos. Hay grupos en los que el cariotipo (conjunto de cromosomas de una célula somática típica de una especie, de un individuo o de una cepa celular determinada) ha sufrido importantes cambios durante el proceso evolutivo, mientras que en otros la dotación cromosómica se muestra extraordinariamente estable, hasta el punto de que es posible que, en estos casos, los procesos de especiación se estén produciendo exclusivamente a través de mutaciones a nivel de los genes.

Por otra parte, mientras que existen cromosomas que han sufrido numerosos e importantes cambios estructurales, otros se hallan, por alguna razón, tan protegidos que se han mantenido invariables durante muchos millones de años.

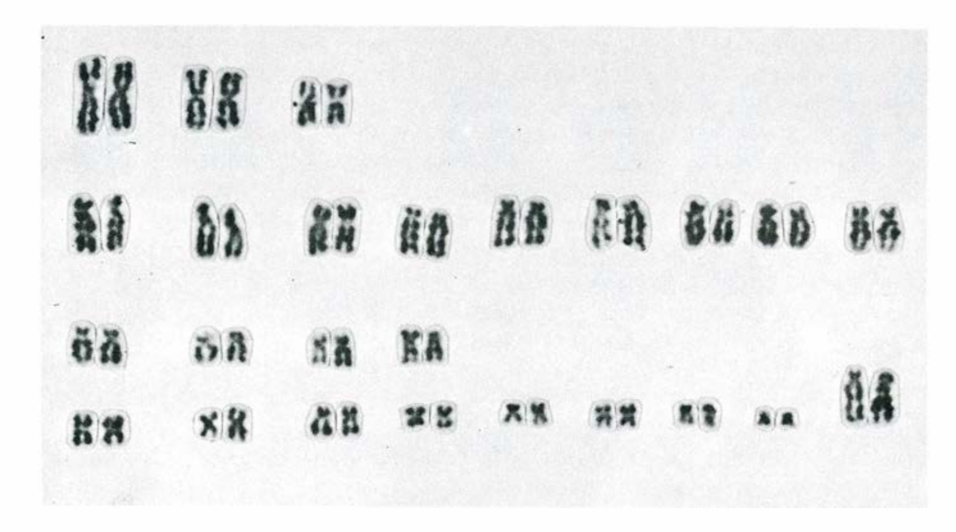
A continuación examinaremos los cambios cromosómicos que se han producido en los diversos grupos de primates, prestando especial atención a aquellas variantes cromosómicas que, por sus características o distribución geográfica, presentan un mayor interés. Como se sabe, el orden de los primates se suele dividir en tres grandes subórdenes: prosimios, platirrinos y catarrinos. Los prosimios abarcan, a su vez, tres grandes grupos: lemuriformes, lorisiformes y tarsiiformes. Los platirrinos comprenden los monos arañas, sapajúes, monos aulladores, tities, monos ardilla, etc. Entre los catarrinos se hallan los macacos, mandriles, manzabeys, cercopitecos, gibones, orangutanes, gorilas, chimpancés y el hombre. La zoología de mamíferos superiores y la paleoantropología se ocupan de justificar la razón de esa clasificación convencional. A la genética compete buscar el fundamento cromosómico de las diferencias fenotípicas y explicar desde ese depurado nivel la dependencia, la filogenia y la evolución.

Los prosimios son los primates que se consideran más primitivos, por más que el concepto de primitividad sea discutible y de difícil aplicación, y arrastre consigo no pocas connotaciones antropocéntricas.

Los prosimios se hallan distribuidos en Africa y Asia. Entre ellos, los lemu-



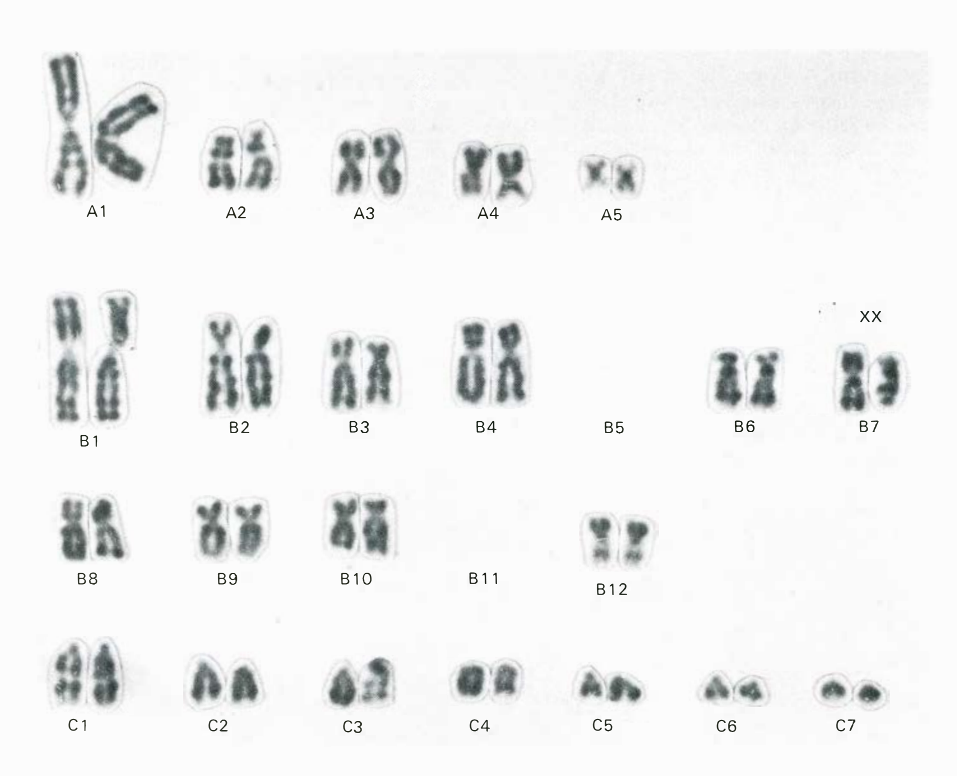
EL LORIS LENTO, *Nycticebus coucang,* **que habita en el Sudeste de Asia, ha sido uno de los prosimios mejor estudiados desde el punto de vista cromosómico. El parecido de su cariotipo con el de otro Lorísido,** *Arctocebus calabarensis,* **que se distribuye en Africa, constituye uno de los más interesantes problemas de evolución cromosómica en primates, problema todavía no resuelto.**



CARIOTIPO DEL LORIS LENTO, *Nycticebus cougang,* **tratado con soluciones salinas calientes para conseguir la aparición de bandas G (véase texto). Esta especie presenta una considerable variabilidad cromosómica, conociéndose individuos con 52, 51 y 50 cromosomas. Algunos de sus pares cromosómicos presentan un patrón de bandas idéntico al de ciertos cromosomas humanos.**



EL MONO ARDILLA, *Saimiri sciureus,* **se encuentra ampliamente distribuído en América del Sur. Por sus características, es una de las especies más utilizadas en investigación biomédica. Gracias a ello, sus cromosomas han sido especialmente bien estudiados. Hasta el momento se han detectado tres tipos cromosómicos, correspondientes a los grupos de Perú, Colombia y la Guayana.**



CARIOTIPO DEL MONO ARDILLA, *Saimiri sciureus,* **tratado con soluciones salinas calientes para obtener bandas G (véase texto). Este tipo cromosómico corresponde al de los animales procedentes de Guayana. En el cariotipo existen varios pares cromosómicos con patrones de bandas idénticos a los de ciertos humanos. M. García y R. Miró estudiaron dicho cariotipo en el laboratorio del autor.**

roideos merecen especial atención, ya que al hallarse distribuídos exclusivamente en Madagascar han permanecido aislados durante un largo período de tiempo, sin posibilidades de interacción con otros grupos.

En los lemuroideos, los números cromosómicos o números diploides oscilan entre 20 y 66. Estas diferencias, con algunas excepciones, se deben a la reducción progresiva del número diploide a través de fusiones céntricas, mecanismo que, como hemos indicado anteriormente, da lugar a una disminución del número cromosómico en dos unidades cada vez que se produce. Las diferencias existentes entre las cifras anteriormente citadas indican de forma expresiva que el número de fusiones céntricas que se ha producido en este grupo ha sido extraordinariamente elevado. Sin embargo, los datos disponibles acerca de este grupo de primates son aun escasos, por lo que no es posible, en estos momentos, confirmar en su totalidad las predicciones con respecto a los cambios cromosómicos que han dado origen a los cariotipos que se observan actualmente.

Aparte de los lemuroideos, la única especie de prosimios a la que se han aplicado hasta el momento las nuevas técnicas de bandas ha sido el loris lento, *Nycticebus coucang,* perteneciente a la subfamilia Lorisinae, cuyos cromosomas han sido estudiados en nuestro laboratorio por M. García y R. Miró. Al no disponerse de otros datos, resulta imposible por el momento comparar el patrón de bandas de *Nycticebus coucang* con el de otros Lorisinae, y aún con otros miembros de su misma especie, en la que se han descrito números cromosómicos de 2n=52 y de 2n=50, correspondientes a individuos homocigotos para el tipo cromosómico antiguo (2n=52) y para el tipo nuevo producido por fusión (2n=50), respectivamente.

Sin embargo, hay que señalar que el cariotipo de *Nycticebus* presenta varios pares cromosómicos con un patrón de bandas idéntico al de ciertos pares de cromosomas de la especie humana, lo que constituye un buen ejemplo de la constancia de dichos patrones a lo largo del proceso evolutivo. Es de esperar que las homologías citadas se confirmen más adelante, a nivel génico, mediante métodos de hibridación celular.

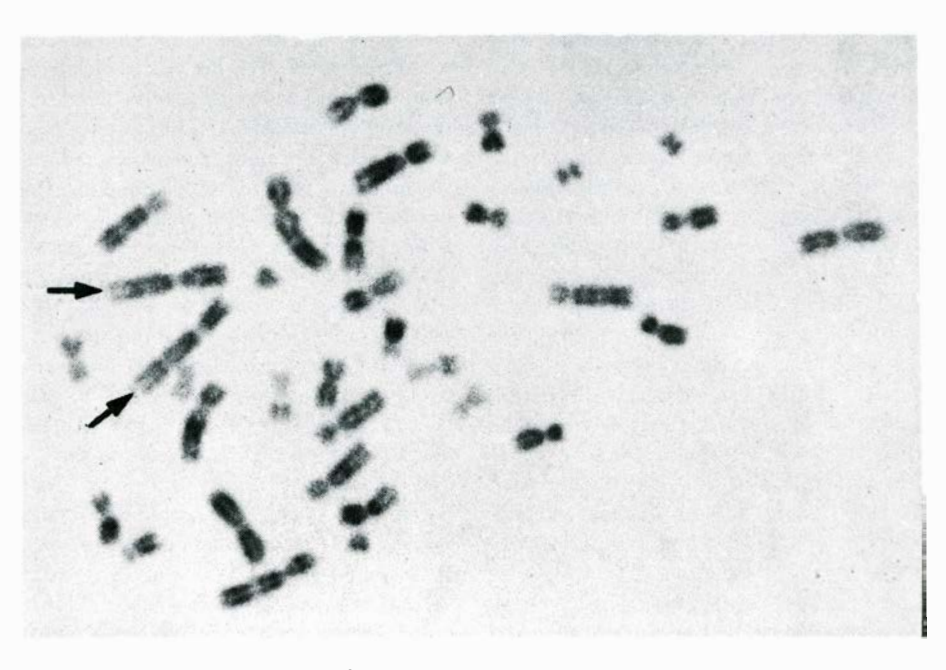
Será interesante la comparación de los patrones de bandas de *Nycticebus coucang* prosimio que habita el S.E. de Asia, con los de *Arctocebus calabarensis,* lorísido del Africa Oriental, cuyo cariotipo es, a grandes rasgos, idéntico al de *Nycticebus.*

De esta forma será posible determinar si se trata de un caso de evolución paralela, que ha dado lugar a una similitud morfológica tan sólo aparente, o si efectivamente, a pesar del tiempo transcurrido desde su separación, estas dos especies presentan aun un cariotipo que no ha sufrido cambios morfológicos apreciables.

Los platirrinos son primates que se hallan distribuidos por América Central y por América del Sur. Este suborden comprende un número muy elevado de especies, de características anatómicas muy variables, correspondientes a tres subfamilias: Cebidae, Callimiconidae y Callithricidae.

Constituyen quizás el grupo más interesante del Orden Primates desde el punto de vista cromosómico, ya que en ellos es en donde los mecanismos de especiación se están produciendo con mayor intensidad.

Al igual que en *Nycticebus coucang*, existen varias especies de platirrinos en las que se han detectado los tres tipos cromosómicos que comportan tanto las fusiones céntricas como las inversiones pericéntricas. Así, por ejemplo, en el mono lechuza, *Aotus trivirgatus*, se conocen seis tipos cromosómicos: la forma "antigua", con un número diploide de 2n=54, los heterocigotos para una, dos y tres fusiones céntricas, con números cromosómicos de 2n=53, 51 y 49 y los homocigotos para los tipos "nuevos" con números diploides de 2n=52 y 50.



METAFASE DEL MONO ARAÑA, *Ateles geoffroyi*, **teñida con dihidrocloruro de quinacrina y observada en un microscopio de fluorescencia para obtener bandas Q (véase texto). El cariotipo consta de 34 cromosomas. El par de mayor tamaño presenta una inversión pericéntrica en heterocigosis, que probablemente provocará problemas reproductivos en este animal, del suborden de los platirrinos.**

En el mono ardilla, *Saimiri sciureus*, estudiado en nuestro laboratorio por M. García y R. Miró, se han detectado dos inversiones pericéntricas, de distribución geográfica bien definida, habiéndose observado también homocigotos para los tipos "antiguo" y "nuevo", y los heterocigotos correspondientes. Lo mismo ocurre en el mono aullador *Alouatta seniculus* y en el mono araña *Ateles geoffroyi* con respecto a una inversión pericéntrica. En el género *Cebus* se han descrito también diversas inversiones pericéntricas de características semejantes.

Por otra parte, se cuenta cada vez con más datos que indican que en los primates americanos no es rara la presencia de más de una línea celular o cariotipo específico, con números y constituciones cromosómicas distintas, cuya presencia no puede explicarse ni a través de un quimerismo (líneas celulares diversas de distinto origen, como ocurre debido al intercambio placentario de células entre gemelos dicigóticos) ni a través de un mosaicismo (líneas celulares distintas de un mismo origen, como resultado de la segregación anormal de los cromosomas en la meiosis).

El significado de este fenómeno no está aun aclarado, ni existen indicios que permitan su interpretación. Sin embargo, todos estos datos ponen de manifiesto que los primates americanos se hallan en un activo período de especiación, lo que probablemente obligará a plantear de nuevo la clasificación de este suborden en un futuro próximo.

Además, en los platirrinos se han descrito dos fenómenos que revisten importancia desde el punto de vista cromosómico.

Por una parte, se trata del único grupo de primates en el que se ha detectado la existencia de cromosomas supernumerarios. Los cromosomas supernumerarios, bien estudiados en las plantas y también en ciertos tipos de animales, son cromosomas de pequeño tamaño cuyo número varía no sólo en diferentes individuos de una misma especie, sino también en distintas células de un mismo individuo. Estos cromosomas, también llamados cromosomas B, no afectan al fenotipo del individuo, aunque pueden ocasionar una disminución de la fertilidad, cuando su número es excesivamente elevado, y provocar problemas en la mitosis y en la meiosis. Su presencia ha sido detectada en *Alouatta seniculus*, mono aullador, en el que el número de cromosomas supernumerarios oscila entre un máximo de cinco y un mínimo de tres, dando así origen a números cromosómicos de 2n=44, 2n=43 y 2n=42.

Ciertos platirrinos presentan también un sistema poco frecuente de determinación sexual. En general, en los primates el macho es heterogamético, es decir, que posee cromosomas sexuales de morfología distinta, que se designan como X e Y, según vimos. La hembra, homogamética, presenta cromosomas sexuales iguales, que se designan como XX. El sistema de determinación sexual normal es, pues, XX/XY.

Sin embargo, tanto en *Callimico goeldii* como en *Alouatta palliata* se han descrito casos en que uno de los cromosomas sexuales, concretamente el cromosoma Y, se hallaba translocado a un autosoma, es decir, se había unido a un cromosoma no sexual. Estos casos se han observado como correspondientes a un sistema de determinación sexual XX/XO, aunque, en propiedad, se trata del sistema de determinación sexual conocido como $X_1X_1X_2X_2/X_1X_2Y$, o mejor, indicando al autosoma que interviene en la translocación con el Y como AY, al autosoma no translocado como A y a los cromosomas X como X, a un sistema XXAA/XAAY. Como podía intuirse, translocación es la rotura de un fragmento de cromosoma que se une a otro cromosoma.

En los catarrinos se distinguen, según los criterios más actuales, cuatro familias: Cercopithecidae, Hylobatidae, Pongidae y Hominidae. Los Cercopitheci-

dae, que incluyen tres subfamilias, Psynopithecinae, Cercopithecinae y Colobinae se hallan distribuidos en Europa (macacos de Gibraltar), Asia y Africa.

Desde el punto de vista cromosómico existen en esta familia grandes diferencias. Por una parte la subfamilia Psynopithecinae, que incluye a los macacos, a los babuínos y a los mangabeys, presenta una uniformidad cromosómica absoluta, como han demostrado en nuestro laboratorio A. Rubio y R. Miró. Hasta el momento, todas las especies estudiadas muestran el mismo cariotipo, y también un patrón de bandas idéntico *(véase la ilustración de la página 79).* Esta identidad se ha comprobado en los tres géneros que constituyen esta subfamilia: *Macaca, Papio* y *Cercocebus.*

Estos datos son interesantes, ya que, a diferencia de lo que ocurre en otros grupos de primates, en los que la variabilidad cromosómica es considerable, o incluso extrema, en los Psynopithecinae no se han producido cambios estructurales a ningún nivel. De hecho, la clasificación de estos géneros y especies de primates ha sido discutida en numerosas ocasiones, con una clara tendencia a reducir su número, ya que, con frecuencia, se producen de forma espontánea híbridos fértiles entre ellas. Como resultado, se ha recurrido a utilizar términos del tipo de semi-especies o de co-especificidad, que nada resuelven, pero que sí indican que, en este grupo, es probable que los procesos de especiación no estén, ni con mucho, completos, no habiéndose establecido todavía las barreras de fertilidad necesarias.

Por otra parte, es también probable que en los Psynopithecinae los procesos de especiación se estén produciendo exclusivamente a través de mutaciones génicas, cuyo efecto sería quizá más lento. Las mutaciones a nivel génico son cambios debidos a una lectura errónea de uno o más codones (triplete de nucleótidos que constituye la unidad de código del ADN), lo que comporta la aparición de un aminoácido en vez de otro.

El grado de homología entre los patrones de bandas de los cromosomas de los macacos, babuínos y mangabeys y los de la especie humana resulta impresionante.

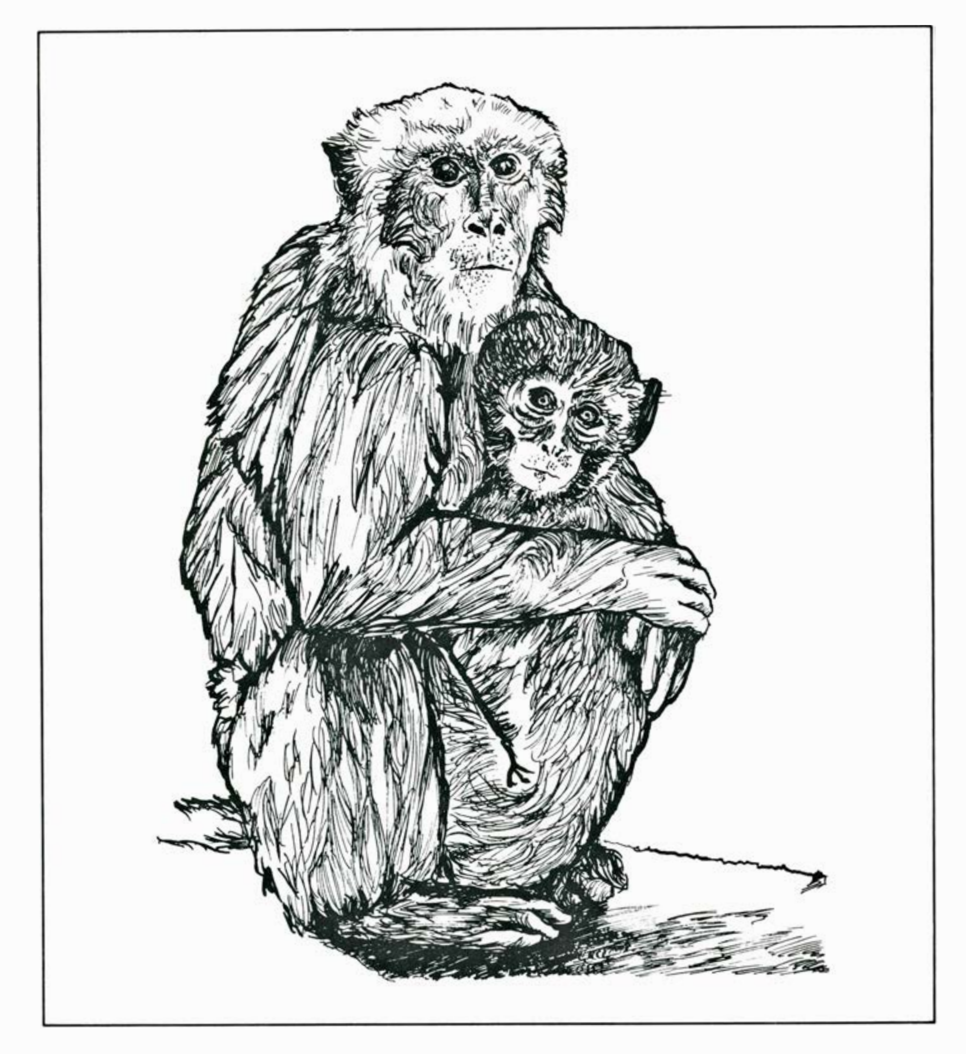
Si se comparan los cariotipos, puede así observarse que el cromosoma número 1 de *Macaca* (o de cualquiera de los otros dos géneros de Psynopithecinae) es comparable al cromosoma 1 humano; el cromosoma 2 de *Macaca,* con una inversión pericéntrica, es comparable al par 3 humano; el cromosoma 3 de *Macaca* al par 6 humano; el par 4 de *Macaca* al par 4 humano; el par 6 de *Macaca* al par 5 humano; el par 7 de *Macaca* al par 7 humano (que en la ilustración pertinente aparece en la posición 8); el par 9 de *Macaca* al par 12 humano; el par 10 de *Macaca* al par 10 humano; el par 11 de *Macaca* al par 8 humano (que en la ilustración mencionada se halla en posición 7); el par 13 de *Macaca* al par 18 humano; el par 15 de *Macaca* al par 11 humano; el par 19 de *Macaca al* par 19 humano; el par marcador, indicado como D del macaco, al par 15 humano; y, finalmente, el cromosoma X de *Macaca,* idéntico al cromosoma X humano.

Esta notable homología constituye un dato más en favor de que los macacos y los babuinos constituyan excelentes modelos para el estudio de las enfermedades de origen genético que afectan al hombre.

En los Cercopithecinae, en cambio, la variabilidad cromosómica es muy considerable. Los números cromosómicos oscilan entre un mínimo de 54 y un máximo de 72, y la morfología de los cromosomas es también extraordinariamente variable. Resulta curioso que, en grupos filogenéticamente tan próximos como son los Psynopithecinae y los Cercopithecinae, la situación a nivel cromosómico resulte tan distinta.

Sin embargo, en los escasos animales de este grupo en los que se ha podido determinar los patrones de bandas, se ha logrado establecer una fuerte correspondencia con los patrones de bandas de los Psynopithecinae. Así, el mono verde africano *Cercopithecus aethiops* y el macaco rhesus *Macaca mulatta* difieren tan sólo en ocho fusiones de tipo céntrico.

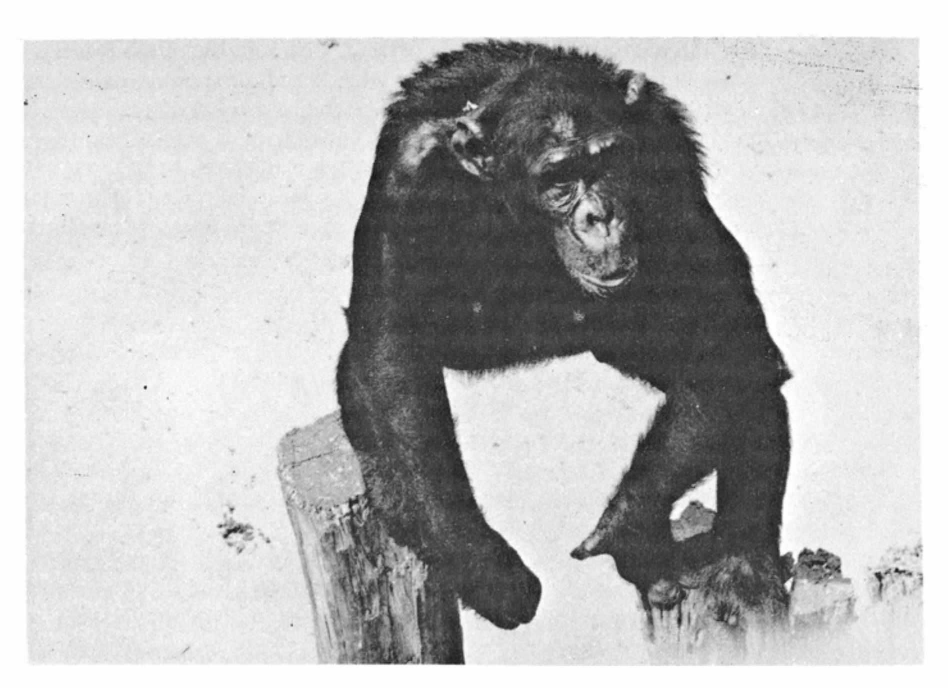
En consecuencia, la homología de los cromosomas de los Cercopithecinae con los cromosomas humanos puede compararse con la indicada para los Psynopithecinae tal y como se ha demostrado hasta el



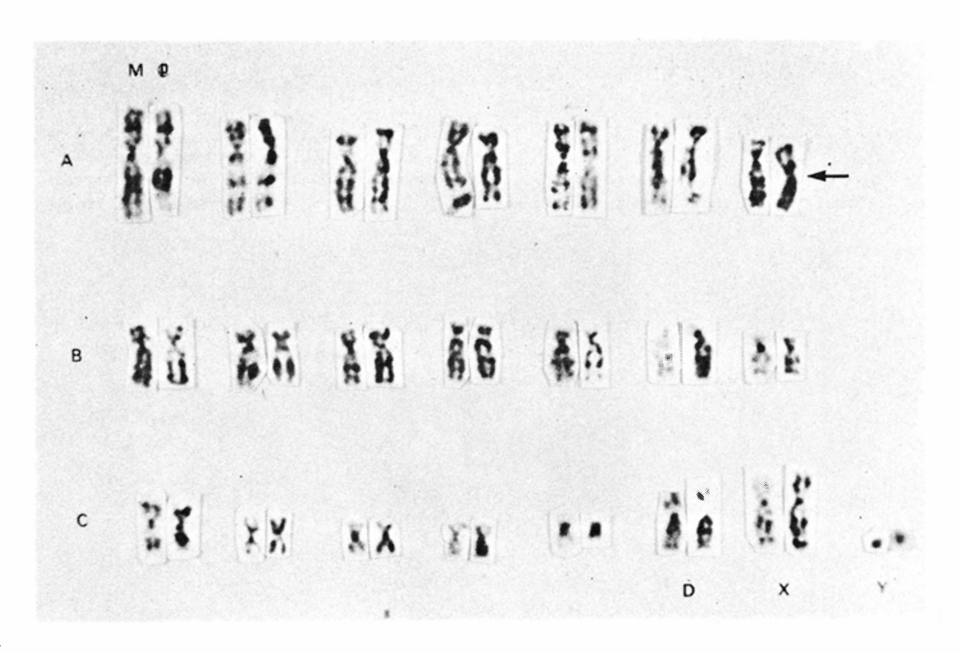
MACACO RHESUS O *MACACA MULATTA*. **Este ubícuo animal, que se encuentra en numerosos pueblos de la India, goza de justa fama gracias al factor sanguíneo que lleva su nombre. Se trata de una especie muy resistente, que se reproduce muy bien en cautividad, y que constituye un excelente candidato como modelo para el estudio de las enfermedades de origen genético que afectan a la especie humana. Hoy se está avanzando mucho en el estudio de su mapa génico.**

momento en las tres únicas especies de este grupo que han sido estudiadas: *Cercopithecus aethiops, Cercopithecus talapoin* y *Cercopithecus nigroviridis.*

Desde el punto de vista evolutivo, y dada la escasez de datos de que se dispone resulta aun imposible determinar con exactitud la situación de los miembros de esta subfamilia.



EL CHIMPANCE, *PAN TROGLODYTES,* es un antropoide africano íntimamente emparentado con el gorila, y también con el hombre. Tanto sus proteínas como las bandas de sus cromosomas presentan una coincidencia de un 99 % con la especie humana. Existen estudios recientes que indican que la capacidad intelectiva del chimpancé es aun mucho mayor de lo que se había imaginado. Más que de un animal "primitivo", se trata de un ser que no ha sido aun comprendido por su congénere (utilizando la palabra con toda intención) humano, pero se tiene conciencia de ello.



COMPARACION ENTRE LOS PATRONES DE BANDAS de *Macaca* y *Papio.* En cada par, el cromosoma de la izquierda corresponde a *Macaca,* y el de la derecha a *Papio* (M y P, respectivamente). Obsérvese que las bandas coinciden en su casi totalidad. La flecha señala la presencia de una inversión pericéntrica, probablemente de tipo esporádico, cuya naturaleza se explicó antes.

En los Colobinae, y desde el punto de vista morfológico, los cariotipos resultan semejantes a los descritos en los Psynopithecinae, pero hasta el momento no se han llevado a cabo en esta subfamilia estudios de bandas que permitan confirmar o descartar esta suposición.

En los Hylobatinae, tal y como se había previsto a partir de datos morfológicos, los cariotipos de *Hylobates lar* y de *Hylobates moloch* han resultado prácticamente idénticos, y muy parecidos a los descritos en los Psynopithecinae. Al igual que en éstos, muchos de los cromosomas presentan un patrón de bandas idéntico al de determinados cromosomas humanos. En cambio, *Hylobates concolor,* con un número cromosómico más elevado, presenta grandes diferencias con los anteriores.

Finalmente, los Pongidae, que incluyen a los orangutanes, gorilas y chimpancés han sido el grupo mejor estudiado desde el punto de vista cromosómico, probablemente a causa del interés que despierta su parecido con la especie humana.

El orangután *(Pongo pygmaeus)* es el que mayores diferencias presenta con respecto al cariotipo de *Homo sapiens.* Estas diferencias, en número de unas 12, resultan a veces de difícil interpretación. Su presencia indica que la rama que dio origen a *Pongo* se separó antes que las que originarían a los gorilas y chimpancés por un lado, y la que, por otro, daría origen a *Homo sapiens.*

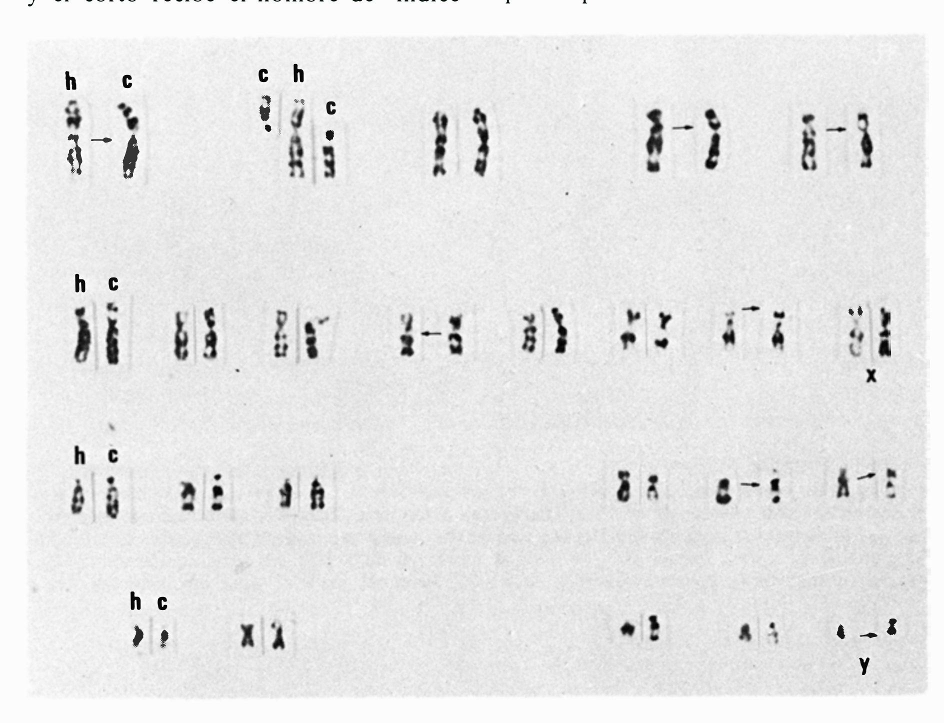
En cambio el chimpancé *(Pan troglodytes)* y el gorila *(Gorilla gorilla)* presentan con la especie humana una identidad de bandas del 99 por ciento, lo que se corresponde con la cifra de identidad a nivel de sus proteínas. Lo mismo puede decirse del chimpancé enano *Pan paniscus.*

Las diferencias estructurales existentes consisten en una fusión céntrica, que dio origen al cromosoma número 2 de la especie humana (de ahí que tanto el chimpancé como el gorila tengan 48 cromosomas, mientras los seres humanos presentan tan sólo 46), y unas ocho inversiones pericéntricas, tal y como ha demostrado en nuestro laboratorio M. R. Caballín. A su vez, chimpancés y gorilas difieren entre sí en dos inversiones pericéntricas, existiendo entre ellos otros dos cambios de difícil interpretación.

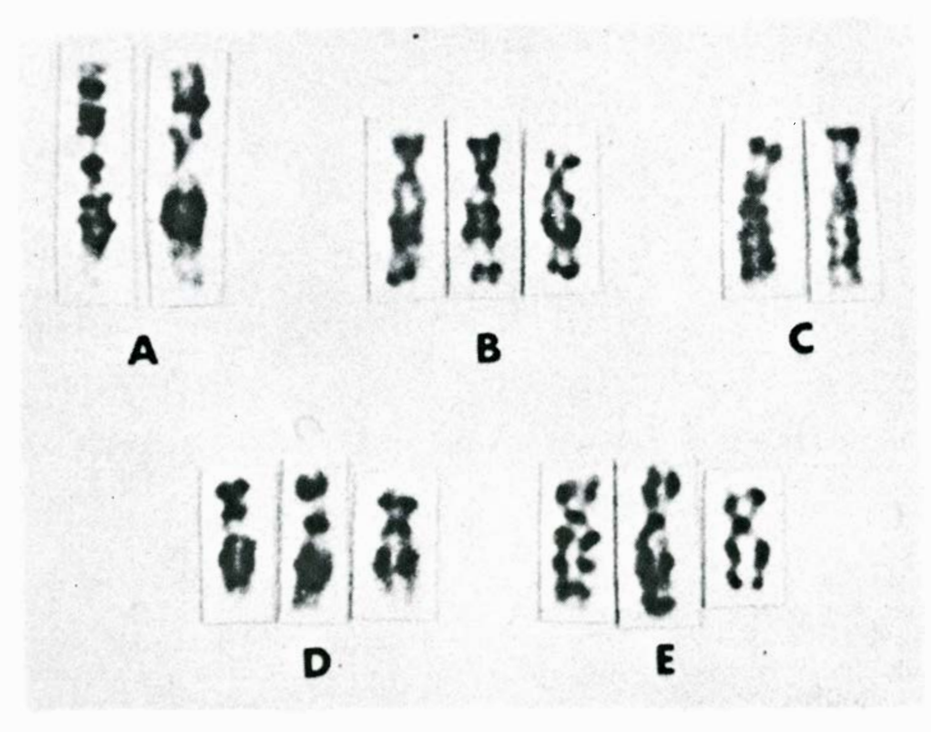
Como puede observarse en la ilustración de la página 80, en la que se comparan los cromosomas del chimpancé con los cromosomas humanos, la identidad de las bandas es prácticamente total, a pesar de la existencia de diversas diferencias estructurales.

Es posible que la fusión céntrica que dio origen al par número 2 humano haya sido el cambio cromosómico que inició el proceso de especiación. Sin embargo, llama también la atención la abundancia de inversiones pericéntricas, cuyo importante papel en la producción de nuevas especies hemos indicado antes.

Hay que señalar también que existen notables diferencias entre el cromosoma humano y el del chimpancé. Una de ellas es la que hace referencia a la forma. Sabido es que la relación o proporción numérica existente entre el brazo largo y el corto recibe el nombre de "índice de brazos" y caracteriza, junto con el tamaño total, a cada cromosoma individual. Cuando el índice de brazos es tan elevado que el centrómero, o punto central de donde parten los brazos, se encuentra en posición casi terminal, a los cromosomas que lo poseen se les llama "acrocéntricos". Cuando el centrómero se encuentra situado en el centro geométrico, el cromosoma es metacéntrico, y submetacéntrico cuando se encuentra alejado, sin llegar a una posición terminal. Pues bien, en *Homo sapiens*, el cromosoma Y es acrocéntrico y posee una amplia zona terminal en sus brazos largos, formada por heterocromatina constitutiva, que es la que presenta una marcada fluorescencia y puede observarse en el núcleo interfásico. Esta zona se utiliza para la determinación del sexo cromosómico, y recibe el nombre de masa Y. En el chimpancé, el cromosoma Y es submetacéntrico, y carece de la zona terminal citada. En cambio, en el gorila el cromosoma Y es prácticamente idéntico al humano.



COMPARACION DE LOS PATRONES DE BANDAS del hombre, *Homo sapiens* **(h) y del chimpancé** *(Pan troglodytes)* **(c). Obsérvese la fusión céntrica que dio origen al cromosoma No. 2 humano. Las flechas señalan otras diferencias estructurales, que, en su mayor parte, son inversiones pericéntricas. La coincidencia de bandas es casi total, como se desprende de su cotejo.**



COMPARACION DE LOS PATRONES DE BANDAS de algunos cromosomas del hombre *(Homo sapiens)*, **del mandril** *(Papio sphinx)* **y del falso tití africano** *(Cercopithecus talapoin)*. **A, cromosoma número 1 humano y equivalente en** *Papio;* **B, cromosoma número 5 humano y equivalente en** *Papio* **y** *Cercopithecus;* **C, cromosoma número 6 humano y equivalente en** *Papio;* **D y E, cromosomas número 12 y X humanos y equivalentes en** *Papio* **y** *Cercopithecus.*

Por supuesto, es falsa la conclusión de que el cariotipo humano constituye el punto final a que han conducido los procesos de evolución cromosómica. A este respecto, existe un interesante estudio realizado por C. Turleau, J. de Grouchy y M. Klein en el que se demuestra que por lo menos cinco de los 23 pares del cariotipo humano corresponderían al tipo primitivo o ancestral, habiendo sufrido menos cambios que los correspondientes cromosomas del chimpancé o del gorila. En general, estos cinco pares de cromosomas presentan características más similares a los del orangután. Asimismo, el gorila y el hombre son los únicos que mantienen la zona altamente fluorescente del cromosoma Y, lo que quizá podría considerarse como un rasgo primitivo, con todas las reservas que ante este concepto hemos expresado.

La organización del material genético en los cromosomas es, pues, extraordinariamente conservadora, habiéndose mantenido constante durante largos períodos de tiempo. Gracias a esta característica resulta aun posible, en la actualidad, identificar cromosomas con patrones de bandas totalmente idénticos en organismos que, desde el punto de vista filogenético, se hallan muy alejados.

La persistencia de los patrones de bandas es tan extraordinaria, que Comings ha llegado a plantear la posibilidad de que, en el linaje que dio origen a los mamíferos, se hubiese producido una duplicación del material genético, es decir, una tetraploidización, basándose en el hecho de que, en los mamíferos actuales, pueden formarse pares de pares, es decir grupos de cuatro cromosomas que presentan una disposición muy parecida de sus bandas.

Desde el punto de vista morfológico hemos indicado ya que la existencia de

patrones de bandas comparables no indicaba necesariamente que existiese una identidad de función.

Sin embargo, al examinar los cromosomas de los Psynopithecinae hemos tenido ocasión de observar que existen por lo menos 14 pares de cromosomas que pueden identificarse con pares cromosómicos humanos. Con un nivel tan elevado de identidad, resulta difícil que el parecido entre los dos grupos pueda atribuirse exclusivamente a la casualidad. Por otra parte, actualmente se está estudiando el mapa génico del macaco rhesus *(Macaca mulatta)* y, por el momento, las homologías de los patrones de bandas han sido confirmadas a nivel génico. Es decir, la identidad morfológica va unida a una identidad funcional, pudiéndose afirmar que las mismas bandas contienen los mismos genes mientras no se demuestre lo contrario.

La posibilidad de identificar a los cromosomas homólogos en especies distintas a través de sus patrones de bandas se pone de manifiesto en la ilustración de la página 80, en la que se comparan algunos pares cromosómicos pertenecientes a especies tan distantes como el falso tití africano *(Cercopithecus talapoin),* el mandril *(Papio sphinx)* y el hombre *(Homo sapiens).* Como puede observarse, la identidad de bandas es prácticamente total.

Los estudios de hibridación celular han sido especialmente amplios en el chimpancé *(Pan troglodytes).* En estos momentos, el mapa génico del chimpancé es prácticamente completo en relación a los *loci* que pueden determinarse en la actualidad. Tal como era de esperar, la localización de los genes presenta una coincidencia absoluta con la situación determinada para los genes humanos. En consecuencia, a nivel cromosómico y a nivel génico nuestro parentesco con el chimpancé debe considerarse como extraordinariamente próximo. Los datos obtenidos en el gorila *(Gorilla gorilla),* aunque todavía sean algo más incompletos, indican la existencia de una situación comparable.

Sin duda, desde el punto de vista evolutivo el valor de los datos que acabamos de exponer es incalculable. Por otra parte, la obtención de mapas génicos en organismos tan próximos al hombre, y que presentan una patología genética equivalente, ofrece una oportunidad extraordinaria para el estudio de la etiología y del posible tratamiento de estos procesos.

Teoría del arco iris

¿Por qué aparecen en el cielo esos bellos arcos coloreados cuando la luz del sol es difundida por gotas de agua? La respuesta a esta difícil pregunta ha puesto a contribución todos los recursos de la física matemática

H. Moysés Nussenzveig

El arco iris es un puente entre dos culturas: tanto para los poetas como para los científicos, su descripción ha venido constituyendo un desafío. Se supone con frecuencia que la descripción científica es un sencillo problema de óptica geométrica, resuelto hace largo tiempo y que hoy conserva interés solamente como ejercicio histórico. No es así: sólo en los últimos años se ha elaborado una teoría cuantitativa satisfactoria del arco iris. Además, esa teoría incluye mucho más que óptica geométrica: se basa en todo lo que sabemos sobre la naturaleza de la luz. Deben tenerse en cuenta las propiedades ondulatorias (la interferencia, la difracción y la polarización) así como las propiedades corpusculares (verbigracia, la cantidad de movimiento transportada por un haz de luz).

Algunas de las más poderosas herramientas de la física matemática se idearon explícitamente para tratar el problema del arco iris y otros estrechamente relacionados con él. De hecho, el arco iris ha servido de piedra de toque para poner a prueba teorías ópticas. Con la más afortunada de esas teorías se puede, hoy, describir el arco iris matemáticamente, es decir, predecir la distribución de la luz en el cielo. Los mismos métodos pueden aplicarse a fenómenos parecidos, tales como los brillantes anillos de color llamados glorias, e incluso a otras clases de arco iris, tales como los atómicos y nucleares.

La comprensión científica de los problemas no siempre ha sido recibida sin reservas. Goethe escribió que el análisis de los colores del arco iris realizado por Newton "baldaría el corazón de la Naturaleza". Un sentimiento análogo fue expresado por Charles Lamb y John Keats; en un banquete celebrado en 1817 propusieron un brindis: "Por la salud de Newton y la ruina de las matemáticas". Sin embargo, los científicos que han contribuido a la teoría del arco iris no fueron en manera alguna insensibles a su belleza. En palabras de Descartes: "El arco iris es una maravilla de la naturaleza tan notable... que yo difícilmente podría haber elegido otro ejemplo mejor para la aplicación de mi método".

El arco único y brillante que se ve tras un chubasco de lluvia o en la rociada de una cascada es el arco iris primario. Ciertamente, su característica más conspicua es su despliegue de colores. Estos varían mucho en brillo y claridad, pero siempre siguen el mismo orden: el violeta es el más interno, mezclándose gradualmente con varios tonos de azul, verde, amarillo y naranja, siendo el rojo el más externo.

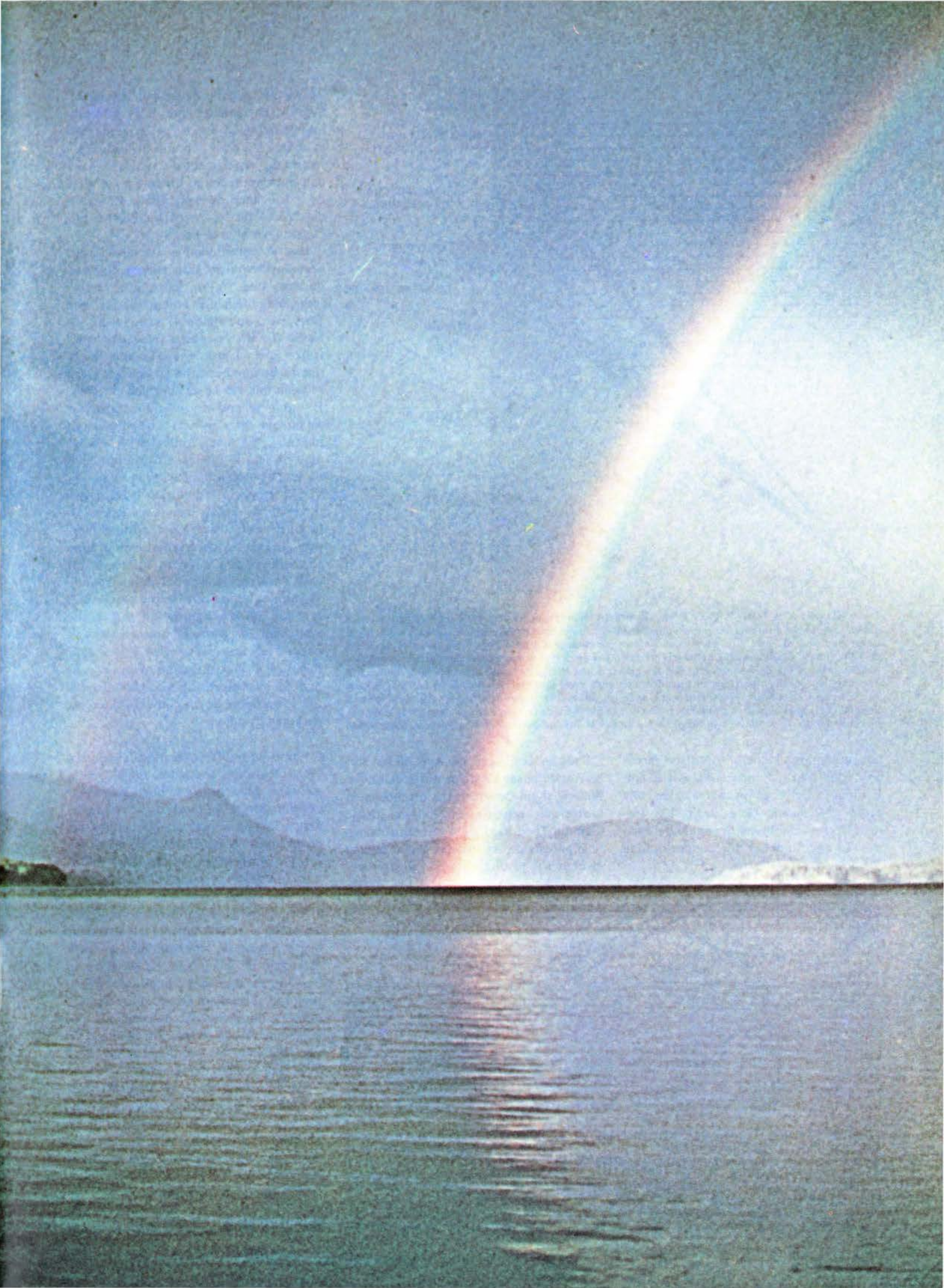
Otras características del arco iris son menos acusadas y, de hecho, no siempre se presentan. En el cielo, por encima del arco primario se encuentra el secundario, en el que los colores aparecen en orden inverso, con el rojo más hacia dentro y el violeta en el borde externo. Una observación cuidadosa revela que la región situada entre ambos arcos es bastante más oscura que el cielo circundante. Incluso en aquellos casos en los que el arco secundario no sea discernible, se puede ver que el arco primario tiene un "lado iluminado" y un "lado oscuro". A la región oscura se la ha denominado banda oscura de Alejandro, en honor del filósofo griego Alejandro de Afrodisias, quien la describió por primera vez hacia el año 200 de la era cristiana.

Otra característica que se ve sólo algunas veces es una serie de débiles bandas, en la que suelen alternar el rosa y el verde, hacia la parte interna del arco primario. (Con menor frecuencia, pueden aparecer en la parte exterior al arco secundario.) Estos "arcos supernumerarios" se ven generalmente con la máxima claridad hacia la parte superior del arco primario. Aunque no son nada llamativos, han tenido una importante influencia en el desarrollo de las teorías sobre el arco iris.

Probablemente, el primero que intentó explicar racionalmente la aparición del arco iris fue Aristóteles. Propuso que el arco iris era en realidad una clase especial de reflexión de la luz del sol por las nubes. La luz se reflejaría a un ángulo fijo, dando lugar a un cono circular de "rayos de arco iris". De esta manera, Aristóteles explicó correctamente la forma circular del arco y advirtió que no se trataba de un objeto material en una posición definida en el cielo, sino más bien de un conjunto de direcciones a lo largo de las cuales se difunde fuertemente la luz hasta los ojos del observador.

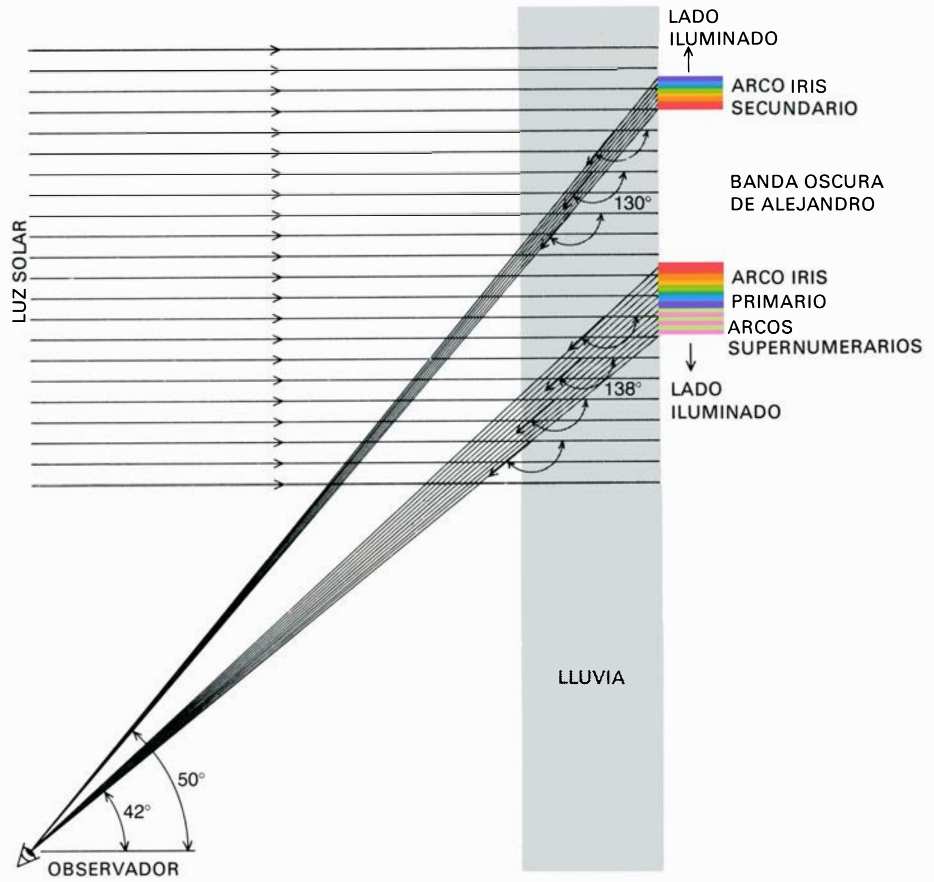
El ángulo que forman los rayos del arco iris y la luz solar incidente fue medido por primera vez, en 1266, por Roger Bacon. Obtuvo un ángulo de unos 42 grados; el arco secundario está unos ocho grados más alto en el cielo. Hoy día, esos ángulos se acostumbran a medir desde la dirección opuesta, de modo

ARCO IRIS DOBLE fotografiado en el Estrecho de Johnstone. La banda brillante más interna es el arco primario; está separado del arco secundario, más débil, por una región, llamada banda oscura de Alejandro, que es visiblemente más oscura que el cielo circundante. Bajo el arco primario hay unas pocas bandas tenues de rosa y verde; son los arcos supernumerarios. A la teoría compete dar una explicación cuantitativa de cada uno de esos rasgos.



lectiva por las gotas de agua en una nube. Sugirió, en cambio, que cada gota era individualmente capaz de producir un arco iris. Además, comprobó esta conjetura en experimentos realizados con una gota aumentada: un frasco esférico de vidrio lleno de agua. Pudo así seguir el camino de los rayos luminosos que constituyen el arco iris.

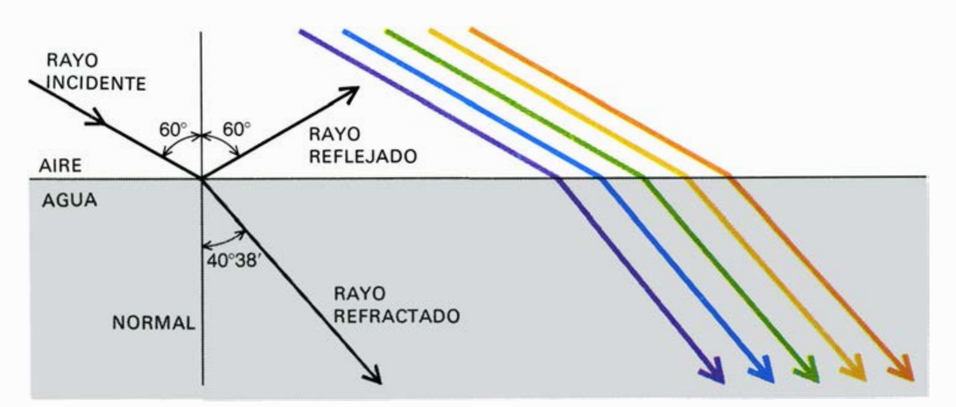
Los descubrimientos de Teodorico permanecieron en gran parte ignorados durante tres siglos, hasta que fueron independientemente redescubiertos por Descartes, que siguió el mismo método. Tanto Teodorico como Descartes mostraron que el arco iris está constituido por rayos que penetran en una gota y se reflejan una vez en su superficie interna. El arco secundario está formado por rayos que han sufrido dos reflexiones internas. En cada reflexión se pierde algo de luz, lo que es la causa principal de que el arco secundario sea más débil que el primario. Teodorico y Descartes también se dieron cuenta de que a lo largo de cada dirección dentro del margen de ángulos correspondiente al arco iris, sólo se podía ver un color a la vez en la luz dispersada por el globo. Cuando se variaba la posición del ojo para explorar los otros ángulos de dispersión, aparecían los otros colores espectrales, de uno en uno. Teodorico y Descartes llegaron a la conclusión de que cada uno de los colores del arco iris llega al ojo procedente de un conjunto diferente de gotas de agua.



LA GEOMETRIA DEL ARCO IRIS viene determinada por el ángulo de desviación que es el ángulo que forma la dirección del rayo incidente con la del rayo emergente de la gota, en el sentido de propagación de éste. Los rayos incidentes se desvían preferentemente según ángulos de 138 y de 130 grados, originando los arcos iris primario y secundario, respectivamente. Entre esos ángulos es muy poca la luz que es devuelta por las gotas; ésa es la región de la banda oscura de Alejandro. Los ángulos óptimos son ligeramente distintos para cada longitud de onda de la luz, por lo cual se dispersan los colores. El orden en que aparecen los colores en el arco secundario es el inverso del que corresponde al arco primario. No existe un plano único en el que se halle el arco iris; éste es el conjunto de direcciones a lo largo de las cuales la luz incidente se desvía hacia el observador.

que medimos el cambio total de dirección de los rayos procedentes del sol. El ángulo del arco primario mide, por tanto, 180 menos 42, es decir, 138 grados; se trata del llamado ángulo del arco iris. El ángulo del arco secundario mide 130 grados.

Tras la conjetura de Aristóteles, hubieron de pasar 17 siglos antes de que se hiciera algún progreso importante en la teoría del arco iris. En 1304, el monje alemán Teodorico de Freiberg rechazó la hipótesis de Aristóteles según la cual el arco iris resultaba de la reflexión co-



LA REFLEXION Y LA REFRACCION de la luz en la superficie de separación entre aire y agua son los hechos básicos en la formación de un arco iris. En la reflexión, el ángulo de incidencia es igual al ángulo de reflexión. En la refracción, el ángulo que la luz transmitida forma con la normal a la superficie de separación viene determinado por las propiedades del medio, caracterizadas por el índice de refraccion. La luz que penetra en un medio con un índice de refracción más alto es desviada acercándose a la normal. La luz de longitudes de onda diferentes se desvía según ángulos ligeramente distintos; esta dependencia entre índice de refracción y color se llama dispersión. Las teorías suelen tratar por separado cada componente monocromática de la luz incidente.

Como ya habían apreciado Teodorico y Descartes, todos los rasgos fundamentales del arco iris se pueden explicar examinando la luz que pasa por una gota única. Los principios fundamentales que determinan la naturaleza del arco son los que rigen la interacción entre la luz y los medios transparentes, es decir, la reflexión y la refracción.

La ley de la reflexión es el principio, conocido e intuitivamente evidente, de que el ángulo de reflexión debe ser igual al ángulo de incidencia. La ley de la refracción es algo más complicada. Mientras que el camino de un rayo reflejado queda completamente determinado por la geometría, la refracción obliga a tener en cuenta las propiedades de la luz y las propiedades del medio.

La velocidad de la luz en el vacío es invariante; en realidad, constituye una de las constantes fundamentales de la naturaleza. Por otra parte, la velocidad de la luz en un medio material queda determinada por las propiedades del medio. El cociente entre velocidad de la luz en el vacío y velocidad de la luz en una substancia se llama índice de refracción de

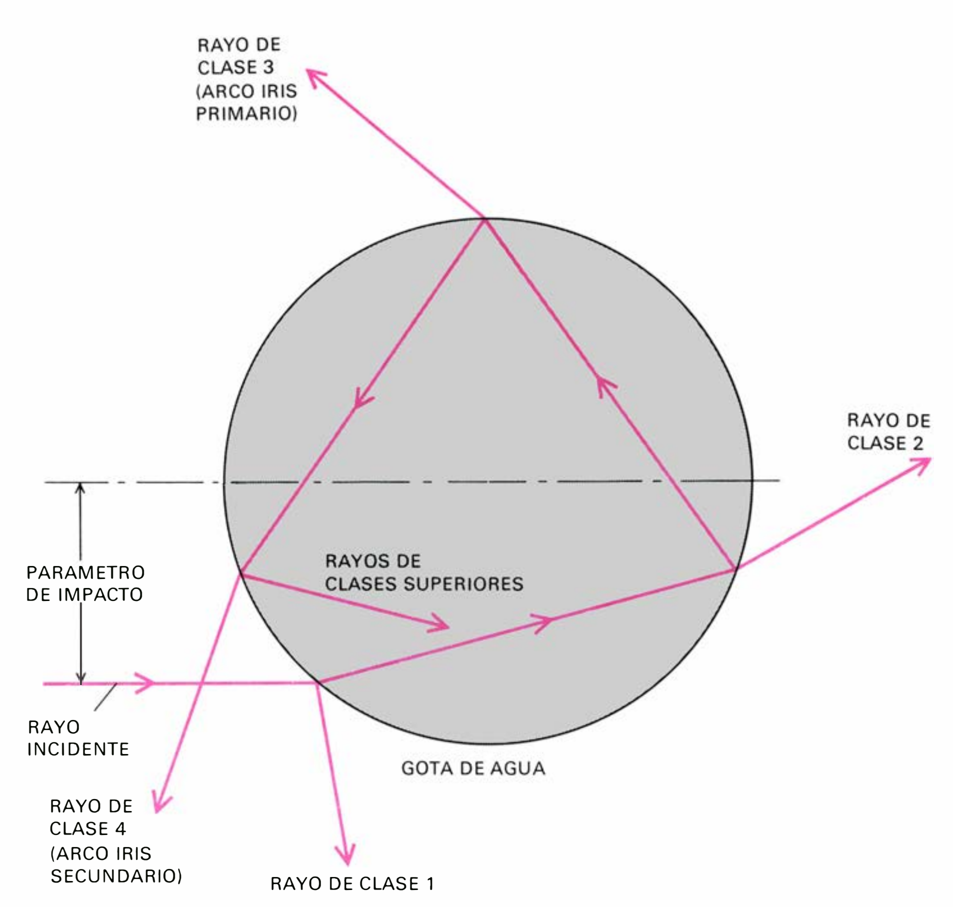
esa substancia. Para el aire, este índice sólo es un poco mayor que la unidad; para el agua vale aproximadamente 1,33.

Un rayo de luz que pasa de aire a agua se retarda al atravesar la superficie de separación; si incide oblicuamente sobre dicha superficie, el cambio de velocidad se traduce en un cambio de dirección. Los senos de los ángulos de incidencia y refracción para dos medios están siempre en una relación constante e igual al cociente de los índices de refracción de los dos medios. Esta igualdad se llama ley de Snell, en honor de Willebrord Snell, quien la formuló en 1621.

Se puede realizar un análisis provisional del arco iris aplicando las leyes de la reflexión y la refracción al camino de un rayo luminoso al atravesar una gota. Admitiendo que la gota es esférica, todas las direcciones son equivalentes y sólo queda una variable importante: la distancia del rayo incidente a un eje paralelo a él y que pase por el centro de la gota. Esta distancia se llama parámetro de impacto, y varía desde cero, cuando el rayo coincide con el eje central, al radio de la gota si el rayo es tangente.

En la superficie de la gota, el rayo incidente es parcialmente reflejado; denominaremos a esta luz reflejada como rayos desviados de clase 1. La luz restante pasa al interior de la gota (con un cambio de dirección causado por la refracción) y en la próxima superficie se transmite de nuevo parcialmente (rayos de clase 2) y se refleja parcialmente. Cuando la luz reflejada vuelve a incidir sobre la superficie, queda separada otra vez entre una componente reflejada y otra transmitida, y el proceso continúa indefinidamente. Así, la gota da origen a una serie de rayos difundidos, cuya intensidad suele decrecer rápidamente. Los rayos de clase 1 representan la reflexión directa por la gota y los de clase 2 se transmiten directamente a través de ella. Los de clase 3 son los que salen de la gota tras una reflexión interna, y forman el arco iris primario. Los rayos de clase 4, que han sufrido dos reflexiones internas, originan el arco secundario. Los arcos de orden superior proceden de rayos con trayectorias más complicadas, pero de ordinario no son visibles.

Para cada clase de rayos desviados, el ángulo de desviación varía dentro de un amplio intervalo de valores en función del parámetro de impacto. Como cuando hay sol la gota queda iluminada simultáneamente para todos los parámetros de impacto, la luz se dispersa prácticamente en todas las direcciones. No es difícil encontrar caminos de rayos luminosos a través de la gota que contribuyen al arco iris, pero hay infinidad de otros caminos que dirigen la luz hacia otro lado. ¿Por qué, entonces, se refuerza la intensidad dispersada en la proximidad del ángulo del arco iris? Esta es una cuestión que Teodorico no examinó; Descartes fue el primero en solucionarla.



EL RECORRIDO DE LA LUZ al atravesar una gota se puede determinar aplicando las leyes de la óptica geométrica. Cada vez que el haz incide sobre la superficie, parte de la luz es reflejada y parte refractada. Los rayos directamente reflejados por la superficie se designan rayos de clase 1; los que se transmiten directamente a través de la gota se llaman de clase 2. Los rayos de clase 3 emergen tras una reflexión interna; éstos son los que originan el arco iris primario. El arco secundario está constituido por rayos de clase 4, que han experimentado dos reflexiones internas. Para rayos de cada clase, el valor del ángulo de desviación está determinado por un factor, el parámetro de impacto (distancia del rayo incidente a un eje paralelo a él y que pase por el centro de la gota).

Aplicando las leyes de la reflexión y refracción en cada punto en el que un rayo incide sobre la superficie de separación aire-agua, Descartes calculó laboriosamente los caminos de muchos rayos incidentes según muchos parámetros de impacto. Los rayos de clase 3 son de importancia predominante. Cuando el parámetro de impacto es cero, los rayos son desviados en la dirección de incidencia, es decir, habiendo pasado por el centro de la gota se reflejan en la pared opuesta de ella, girando un ángulo de 180°, regresando hacia el sol. A medida que crece el parámetro de impacto y los rayos incidentes se separan del centro de la gota, el ángulo de desviación disminuye. Descartes halló, sin embargo, que esta variación no continúa cuando el parámetro de impacto ha aumentado hasta su valor máximo, caso de incidencia rasante en que el rayo es tangente a la superficie de la gota. Por el contrario, el ángulo de desviación pasa por un mínimo cuando el parámetro de impacto vale aproximadamente 7/8 del radio de la gota, y después aumenta de nuevo. El ángulo de desviación correspondiente al mínimo vale 138 grados.

Para los rayos de clase 4, el ángulo de desviación es nulo cuando lo es el parámetro de impacto; en otras palabras, el rayo central se refleja dos veces y después continúa en su dirección original. Al aumentar el parámetro de impacto también lo hace el ángulo de desviación, pero de nuevo esta variación se invierte en cierto momento, esta vez a 130 grados. Los rayos de clase 4 tienen un ángulo de desviación máximo de 130 grados, y si el parámetro de impacto aumenta más, los rayos salen apuntando hacia la parte posterior de la gota, desviados hacia la dirección de incidencia.

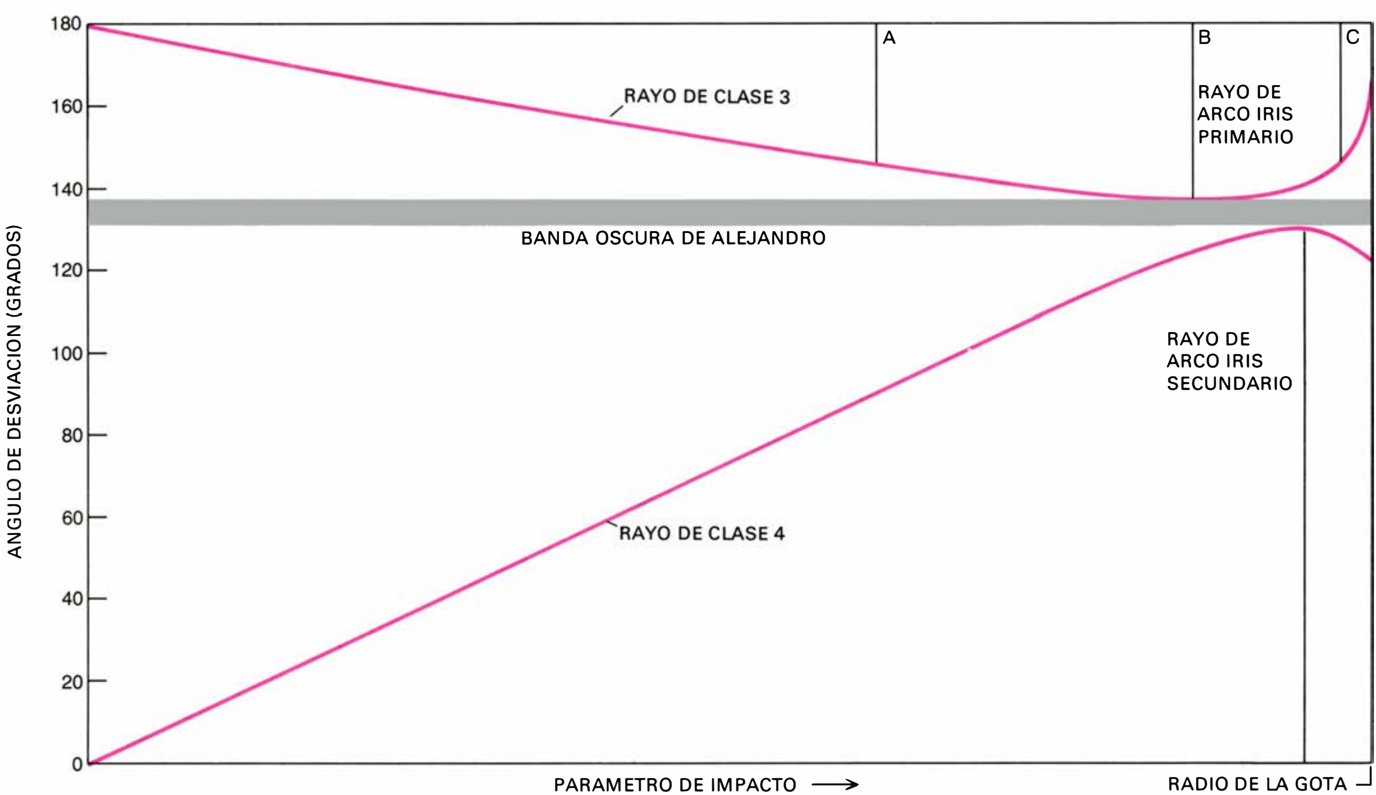
Puesto que, a lo largo del camino del haz solar, la gota es iluminada uniformemente, los parámetros de impacto de los rayos incidentes están también distribuidos uniformemente. Es de esperar, por tanto, que la concentración de luz desviada sea máxima donde más lenta

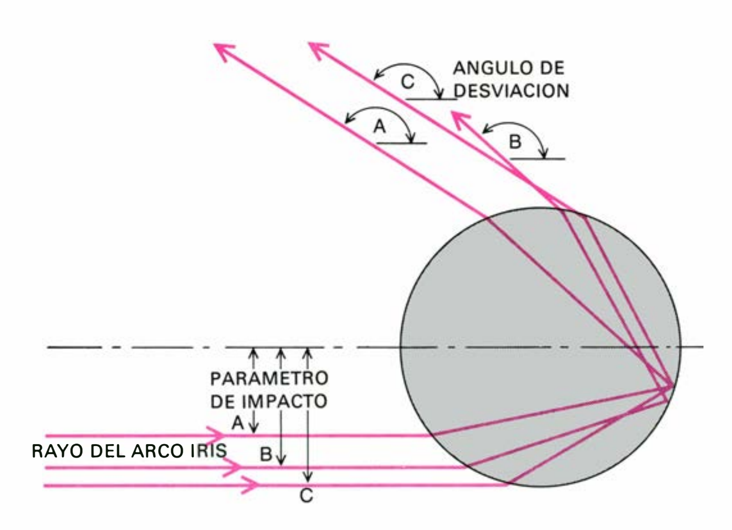
sea la variación del ángulo de desviación al variar el parámetro de impacto. Es decir, la luz desviada es más brillante donde reúne rayos que incidieron con la máxima variedad de parámetros de impacto. Las regiones de mínima variación son las vecinas a los ángulos de máximo y mínimo, y eso explica el especial carácter de los ángulos de los arcos primario y secundario. Además, como en la región comprendida entre los ángulos de 130 y 138 grados no puede haber rayos desviados de las clases 3 y 4, queda también explicada la banda oscura de Alejandro.

La teoría de Descartes se puede entender mejor considerando un conjunto de gotas que, de alguna manera, difundieran la luz uniformemente en todas direcciones. Un cielo lleno de tales gotas sería uniformemente brillante cualquiera que fuese el ángulo. En un cielo poblado de gotas de agua reales hay disponible la misma cantidad de luz, pero ha sido redistribuida. La mayoría de las partes del cielo están iluminadas más débilmente de lo que estarían con dispersión uniforme, pero en la proximidad del ángulo del arco iris hay un arco brillante, que se atenúa gradualmente en el lado iluminado y más bruscamente en el lado oscuro. El arco secundario es una región análoga de alta intensidad, excepto que es más estrecho y todas sus características son más débiles. En la teoría cartesiana, la región entre arcos es decididamente más oscura que en cualquier otra parte: si sólo existieran rayos de las clases 3 y 4, sería muy negra.

El arco iris cartesiano es un fenómeno notablemente sencillo. El brillo es una función de la velocidad de variación del ángulo de desviación. Este queda determinado por dos factores: el índice de refracción, que se supone constante, y el parámetro de impacto, que se supone distribuido uniformemente. Un factor que carece de influencia alguna sobre el ángulo del arco iris es el tamaño de las gotas: la geometría de la desviación es la misma para las gotitas de agua que constituyen las nubes que para los globos de vidrio llenos de agua que usaban Teodorico y Descartes.

Hasta ahora hemos prescindido de una de las características más conspicuas del arco iris: sus colores. Por supuesto, éstos habían sido explicados por Newton en sus experimentos con el prisma en 1666. Tales experimentos demostraron



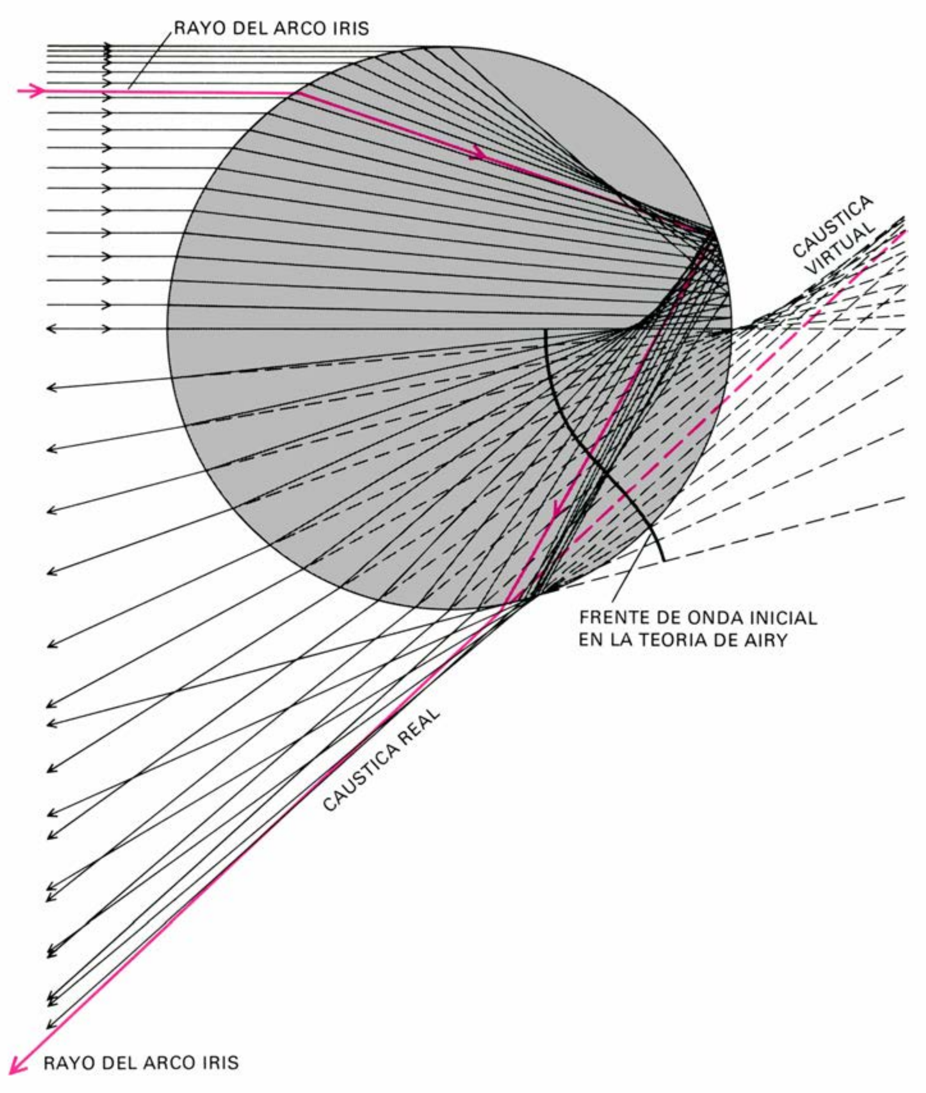


ANGULO DEL ARCO IRIS. La especial importancia que tiene se observa al considerar el ángulo de desviación en función del parámetro de impacto. Cuando éste es cero, el ángulo de desviación para un rayo de clase 3 es 180 grados; el rayo pasa por el centro de la gota, se refleja en la superficie de la gota que encuentra más allá del centro y vuelve hacia el sol en la dirección de incidencia pero en sentido opuesto. A medida que aumenta el parámetro de impacto disminuye el ángulo de desviación, pero llega a alcanzar un valor mínimo. Este rayo de mínima desviación es el rayo del arco iris en el diagrama de la izquierda; los rayos con parámetros de impacto a uno y otro lado de él se desvían según ángulos mayores. La desviación mínima es aproximadamente de 138 grados, y la máxima concentración de rayos desviados se encuentra en la proximidad de este ángulo. El incremento resultante de la intensidad de la luz difundida se percibe como arco primario. El arco secundario se forma de manera análoga, excepto que el ángulo de desviación para los rayos de clase 4 que lo forman aumenta hasta un máximo en lugar de disminuir hasta un mínimo. El máximo se encuentra hacia 130 grados. Ningún rayo de las clases 3 y 4 puede alcanzar ángulos de desviación entre 130 y 138 grados, lo que explica la existencia de la banda oscura de Alejandro. A la izquierda se ve cómo dos rayos de clase 3, con parámetros de impacto a uno y otro lado del correspondiente al rayo del arco iris, emergen según el mismo ángulo de desviación. La interferencia entre rayos como éstos da lugar a los arcos supernumerarios.

no solamente que la luz blanca es una mezcla de colores, sino también que el índice de refracción difiere para cada color, efecto llamado dispersión. Se deduce, por tanto, que cada color o longitud de onda de luz debe tener su propio ángulo de arco iris; lo que observamos en la naturaleza es una reunión de arcos monocromáticos, cada uno ligeramente corrido respecto al anterior.

Partiendo de sus mediciones del índice de refracción, Newton calculó que el ángulo del arco iris es de 137 grados 58 minutos para la luz roja y de 139 grados 43 minutos para la luz violeta. La diferencia entre ambos ángulos es de 1 grado 45 minutos, que debería ser la anchura del arco iris si los rayos de la luz del sol incidente fueran exactamente paralelos. Aceptando medio grado para el diámetro aparente del sol, Newton obtuvo una anchura total de dos grados 15 minutos para el arco primario.

Juntando las aportaciones de Descartes y Newton, quedaban definidas las características más importantes del arco iris. Explicaron la existencia de los arcos primario y secundario y de la zona oscura que los separa. Calcularon la posición angular de esas características y describieron la dispersión de la luz difundida formando un espectro. Todo ello se consiguió usando exclusivamente la óptica geométrica. Su teoría, sin embargo, tenía un fallo importante: era incapaz de explicar los arcos supernumerarios. La comprensión de esos detalles aparentemente pequeños requiere poseer una idea más sutil sobre la naturaleza de la luz.



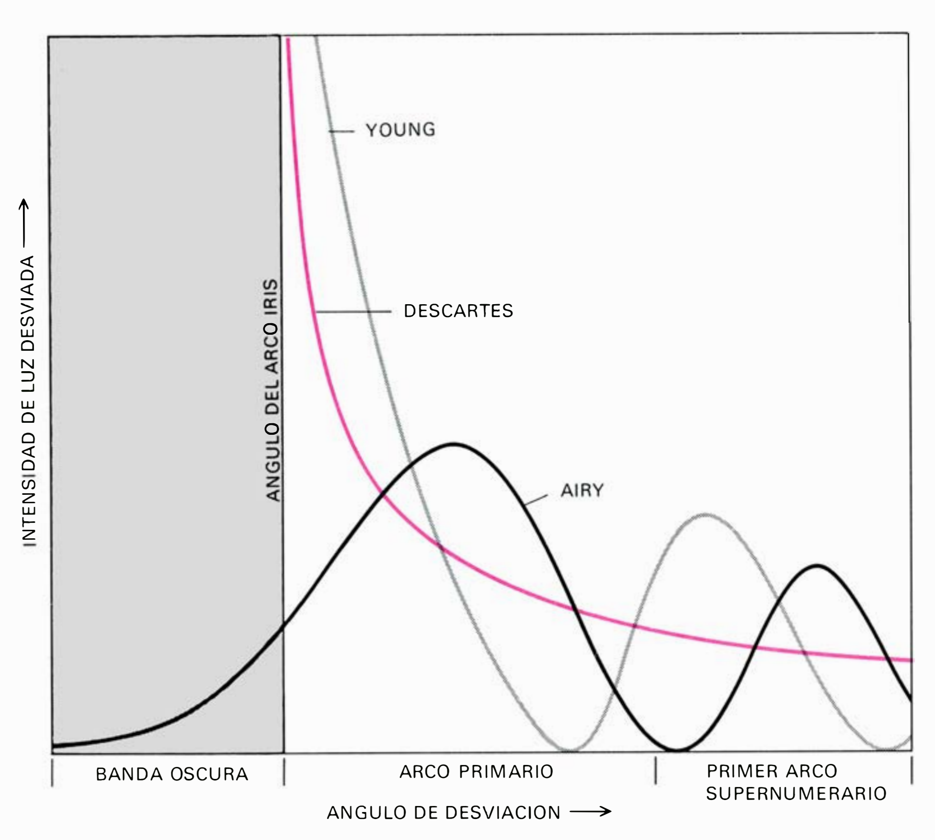
LA CONFLUENCIA DE RAYOS desviados por una gota da lugar a cáusticas. Una cáustica es la envolvente de un sistema de rayos. De interés especial es la cáustica de los rayos de clase 3, que tiene dos ramas, una real y otra virtual; la última se forma por prolongación de los rayos hacia atrás. Cuando se construye el rayo del arco iris en ambas direcciones, se aproxima asintóticamente a las ramas de la cáustica. George B. Airy elaboró una teoría del arco iris basada en el análisis de tal cáustica. Habiendo escogido un frente de onda inicial –una superficie perpendicular en todos los puntos a los rayos de clase 3– Airy pudo determinar la distribución de las amplitudes en las ondas restantes. Un punto débil de esta teoría es la necesidad de presuponer las amplitudes iniciales.

Los arcos supernumerarios aparecen en el lado interno, o iluminado, del arco primario. En esta región angular hay dos rayos de clase 3 que, tras haber sido difundidos, emergen en la misma dirección; provienen de rayos incidentes que tienen parámetros de impacto a uno y otro lado del valor que corresponde al arco iris. Así, para cualquier ángulo dado ligeramente mayor que el del arco iris, la luz difundida incluye rayos que han seguido dos caminos diferentes en el interior de la gota. Los rayos emergen en distintos lugares de la superficie de la gota; siguen en la misma dirección.

En los tiempos de Descartes y Newton, esas dos contribuciones a la intensidad difundida sólo podían manejarse por simple suma. Como resultado, la intensidad que obtenían debía decrecer continuamente con la desviación respecto al ángulo del arco iris, sin traza alguna de arcos supernumerarios. Ahora bien, las intensidades de los dos rayos no se pueden sumar porque no son fuentes independientes de radiación.

El efecto óptico escondido tras el fenómeno de los arcos supernumerarios fue descubierto en 1803 por Thomas Young, quien demostró que la luz era capaz de producir interferencia, fenómeno éste que ya era conocido por el estudio de las olas en el agua. En cualquier medio, la superposición de ondas puede dar lugar bien a un refuerzo (cresta sobre cresta) o bien a una anulación (cresta sobre seno). Young demostró la interferencia de las ondas luminosas haciendo pasar un delgado haz de luz monocromática a través de dos orificios de diámetro muy pequeño y observando las franjas alternativamente iluminadas y oscuras que se producían. El propio Young señaló la relación entre su descubrimiento y los arcos supernumerarios del arco iris. Los dos rayos desviados en la misma dirección por una gota son estrictamente análogos a la luz que pasa a través de los dos agujeros en el experimento de Young. Para ángulos muy próximos al del arco iris, los dos caminos a través de la gota difieren muy poco; por ello los dos rayos interfieren constructivamente. Al aumentar el ángulo, los dos rayos siguen caminos de longitudes substancialmente diferentes. Cuando la diferencia es igual a la mitad de la longitud de onda, la interferencia es completamente destructiva; para ángulos mayores, los dos haces se refuerzan de nuevo. El resultado es una variación periódica en la intensidad de la luz desviada, una serie de bandas alternativamente brillantes y oscuras.

Como los ángulos de desviación a los cuales se produce interferencia cons-



LA INTENSIDAD CALCULADA en función del ángulo de desviación se compara en tres de las primitivas teorías del arco iris. En el análisis geométrico de Descartes, la intensidad es infinita para el ángulo del arco iris; disminuye gradualmente (sin arcos supernumerarios) en el lado iluminado y cae bruscamente a cero en el lado oscuro. La teoría de Thomas Young, basada en la interferencia de las ondas luminosas, predice los arcos supernumerarios, pero retiene la transición brusca de la intensidad de cero a infinito. La teoría de Airy altera la posición de los máximos en la curva de intensidad, y por primera vez proporciona (mediante la difracción) una explicación para la atenuación gradual del arco iris hasta pasar a sombra (diagrama realizado por Gabor Kiss).

tructiva están determinados por la diferencia de longitud entre los caminos de los dos rayos, esos ángulos están afectados por el radio de la gota. El aspecto de los arcos supernumerarios (en contraste con el ángulo del arco iris) depende, pues, del tamaño de las gotas. Para gotas grandes, la diferencia de caminos aumenta mucho más rápidamente con el parámetro de impacto que para gotas pequeñas. De aquí que, cuanto mayores sean las gotas, tanto más estrecha es la separación angular entre los arcos supernumerarios. Los arcos son casi imposibles de distinguir si las gotas son de diámetro superior a aproximadamente un milímetro. La superposición de colores tiende también a eliminar los arcos. La dependencia de los arcos supernumerarios respecto al tamaño de las gotas explica por qué son más fáciles de ver cerca de la cima del arco iris: las gotas de lluvia tienden a hacerse mayores a medida que caen.

Con la teoría de la interferencia de Young se explicaban todas las características importantes del arco iris, al menos en forma cualitativa y aproximada. Faltaba, sin embargo, una teoría matemática, cuantitativa, capaz de predecir la intensidad de la luz desviada como función del tamaño de la gota y del ángulo de desviación.

La explicación de Young de los arcos supernumerarios se basaba en la teoría ondulatoria de la luz. Paradójicamente, sus predicciones para el otro lado del arco, para la banda oscura de Alejandro, estaban en desacuerdo con dicha teoría. La teoría de la interferencia, como las de Descartes y Newton, predecía completa oscuridad en esta región, al menos cuando sólo se consideraban los rayos de las clases 3 y 4. Sin embargo, no es posible una transición tan abrupta porque la teoría ondulatoria de la luz exige que las acusadas separaciones entre luz y sombra se suavicen a causa de la difracción. La manifestación más familiar de la difracción es la manera en que la luz o el sonido parecen curvarse en el borde de un obstáculo opaco. En el arco iris no hay obstáculo real alguno; de todas maneras, el borde de separación entre el arco primario y la banda oscura debería presentar difracción. El tratamiento de la difracción es un sutil y difícil problema de física matemática; el desarrollo subsiguiente de la teoría del arco iris fue estimulado principalmente por los esfuerzos realizados para resolverlo.

En 1835, Richard Potter, de la Universidad de Cambridge, hizo notar que el cruce de varios conjuntos de rayos luminosos en una gota da lugar a curvas cáusticas. Una cáustica, o "curva quemante", representa la envolvente de un sistema de rayos y va siempre asociada a un alto valor de la intensidad. Una cáustica conocida es la curva brillante de dos arcos que se unen en punta y que se forma en una taza de té cuando la luz del sol se refleja en sus paredes interiores. Las cáusticas, como el arco iris, generalmente tienen un lado iluminado y otro oscuro; la intensidad aumenta continuamente hasta la cáustica y, pasada ésta, cae en forma brusca.

Potter mostró que el rayo de Descartes del arco iris –el rayo de clase 3 con ángulo de desviación mínimo– se puede considerar como una cáustica. Todos los otros rayos transmitidos de clase 3, cuando se prolongan hasta el infinito, se acercan al rayo de Descartes desde el lado iluminado; no hay rayos de esta clase en el lado no iluminado. Así, hallar la intensidad de la luz desviada en un arco iris es semejante al problema de determinar la distribución de intensidades en la vecindad de una cáustica.

En 1838, George B. Airy, colega de Potter en Cambridge, intentó determinar tal distribución. Su razonamiento se basaba en un principio de la propagación de ondas formulado en el siglo XVII por Christiaan Huygens y perfeccionado más tarde por Agustín Juan Fresnel. Este principio considera cada punto de un frente de onda como fuente de ondas esféricas secundarias; las ondas secundarias definen un nuevo frente de onda; de esta manera se explica la propagación de las ondas. Se deduce, por tanto, que si se conociera la amplitud de las ondas secundarias en cada punto de un frente de onda completo, se podría reconstruir la distribución de amplitudes en cualquier otro punto. La totalidad del arco iris se podría describir rigurosamente si conociéramos la distribución de amplitudes a lo largo del frente de onda en una gota única. Por desgracia, la distribución de amplitudes raramente puede determinarse; lo más que generalmente se puede hacer es una estima razonable para un frente de onda escogido con la esperanza de obtener una buena aproximación.

El frente de onda de partida escogido por Airy es una superficie en el interior de la gota, normal a todos los rayos de clase 3 y con un punto de inflexión (cambio en el sentido de la curvatura) donde corta al rayo de Descartes. Las amplitudes de las ondas a lo largo de este frente de ondas se estimaron mediante hipótesis corrientes en la teoría de la di-

fracción. Airy pudo entonces expresar la intensidad de la luz desviada en la región del arco iris en términos de una nueva función matemática, entonces conocida como integral del arco iris y hoy llamada función de Airy. No trataremos aquí de la forma matemática de la función de Airy.

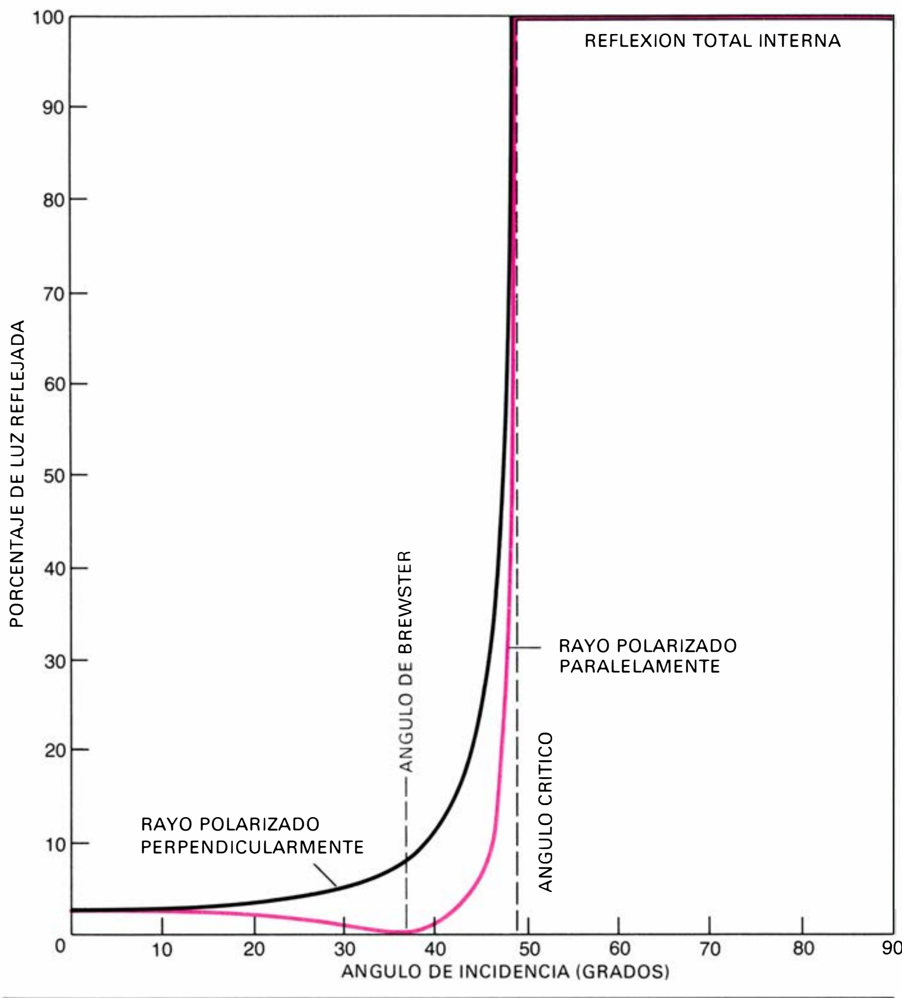
La distribución de intensidades predicha por la función de Airy es análoga a la figura de difracción que aparece en la sombra de un filo rectilíneo. En el lado iluminado del arco primario hay oscilaciones de intensidad que corresponden a los arcos supernumerarios; las posiciones y anchuras de esos máximos difieren algo de lo predicho por la teoría de la interferencia de Young. Otra diferencia importante en la teoría de Airy es que la intensidad máxima del arco iris se produce para un ángulo algo mayor que el ángulo de mínima desviación de Descartes. Las teorías de Descartes y Young predicen una intensidad infinita para ese ángulo a causa de la presencia de la cáustica. En la teoría de Airy no se alcanza intensidad infinita en ningún punto, y para el rayo de Descartes, la intensidad pronosticada por la teoría es menos de la mitad que la que corresponde al máximo. Por último, en el lado oscuro del arco iris aparecen efectos de difracción; en lugar de anularse bruscamente, la intensidad disminuye poco a poco en la banda oscura de Alejandro.

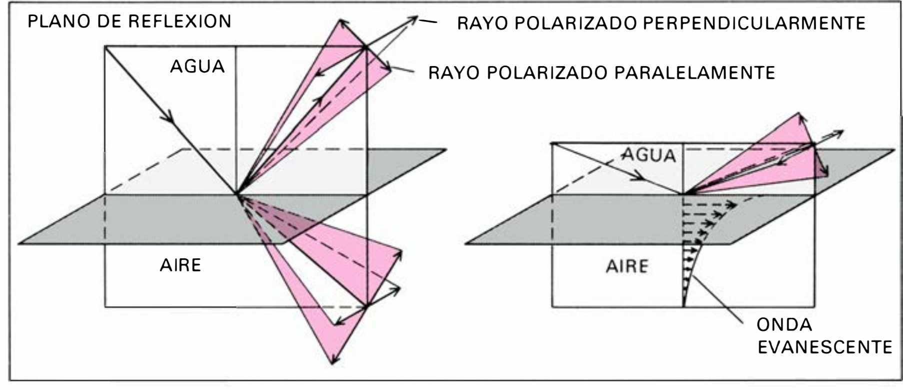
Airy realizó sus cálculos para un arco iris monocromático. Si se quiere aplicar su método al arco iris producido en la luz del sol, hay que superponer las distribuciones de Airy generadas por las diversas componentes monocromáticas. Pasar más allá y describir la imagen del arco iris que se percibe exige una teoría de la visión del color.

La pureza de los colores del arco iris queda determinada por el grado de superposición de los arcos iris monocromáticos componentes, la que a su vez viene determinada por el tamaño de las gotas. Gotas uniformemente grandes (de diámetro de unos cuantos milímetros) dan generalmente arcos iris brillantes con colores puros; con gotas muy pequeñas (diámetros de 0,01 milímetro aproximadamente) la superposición de colores es tan grande que la luz resultante aparece casi blanca.

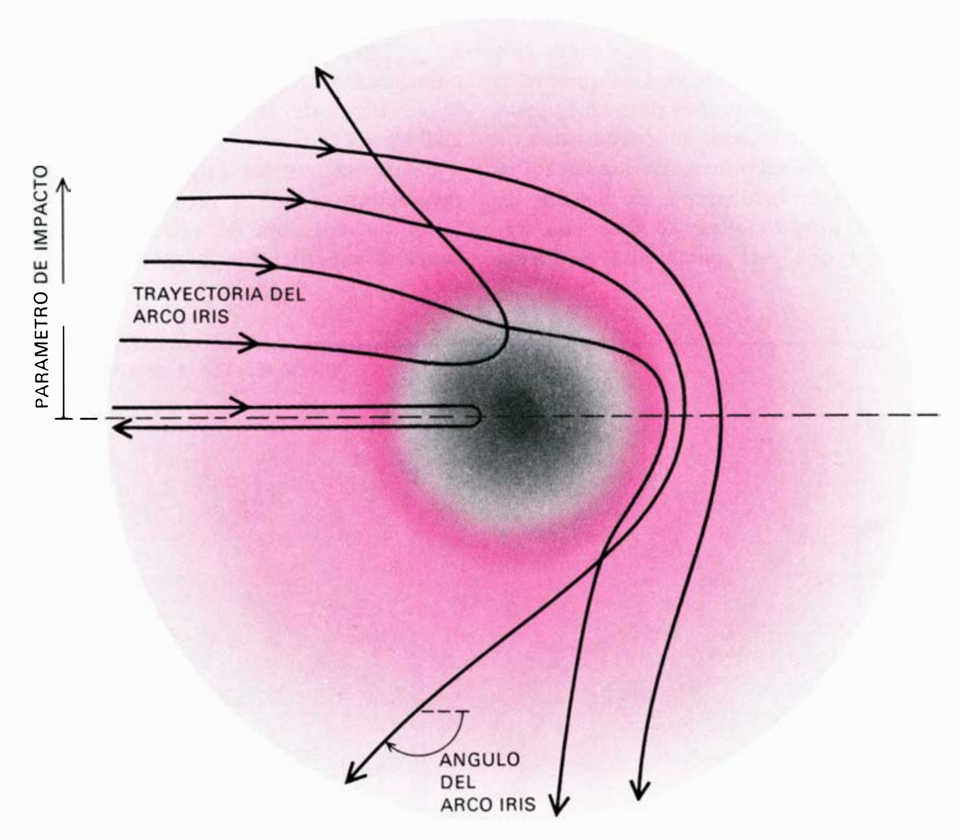
Una propiedad importante de la luz, de la que hasta ahora hemos prescindido, es su grado de polarización. La luz es una onda transversal, es decir, una onda en la que las oscilaciones son perpendiculares a la dirección de propagación. (El sonido, por el contrario, es una vibración longitudinal.) La orientación de la oscilación transversal se puede separar en componentes a lo largo de dos ejes perpendiculares entre sí. Cualquier rayo luminoso se puede describir en función de esos dos estados independientes de polarización rectilínea. La luz del sol es una mezcla incoherente de los dos en iguales proporciones; se dice con frecuencia que está polarizada en forma aleatoria, o simplemente no polarizada. La reflexión puede alterar el estado de polarización, y en ese hecho radica la importancia de la polarización en el análisis del arco iris.

Consideremos las reflexiones de un rayo luminoso que se propaga en el interior de una gota de agua cuando alcanza la superficie de la gota. El plano

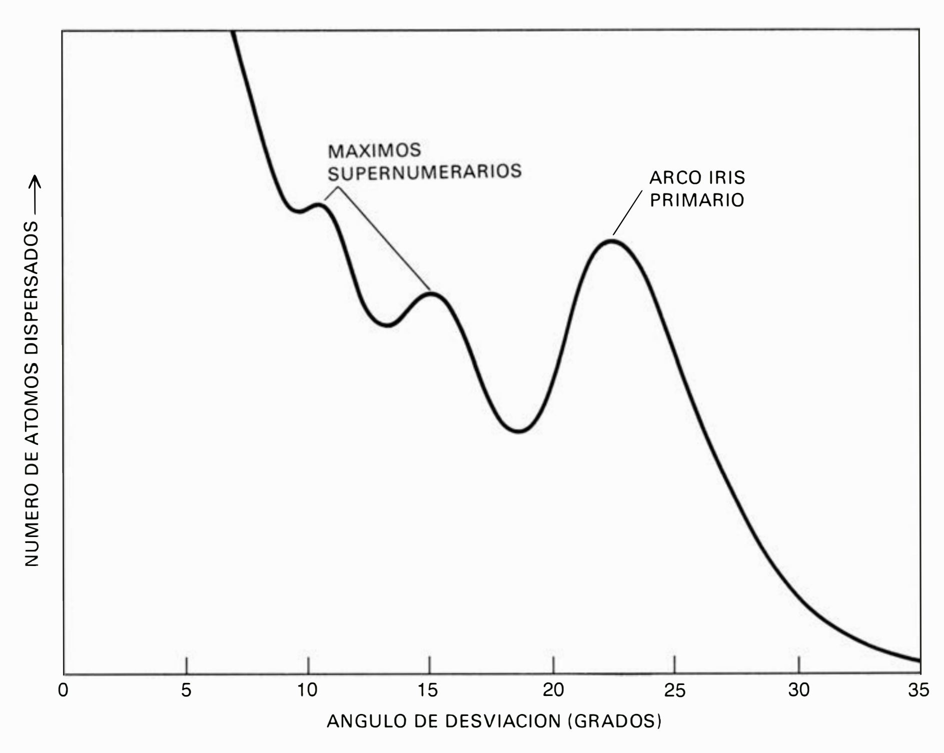




LA POLARIZACION DEL ARCO IRIS es el resultado de la reflexión diferencial. Un rayo incidente se puede separar en dos componentes, una polarizada paralelamente y la otra perpendicularmente al plano de reflexión. Para un rayo que incida sobre la superficie de separación entre aire y agua desde dentro de una gota, la reflectividad de la superficie depende del ángulo de incidencia. Mas allá de un ángulo crítico, ambas componentes, la perpendicular y la paralela, son totalmente reflejadas aunque algo de luz viaja paralelamente a la superficie constituyendo una "onda evanescente". Para ángulos menores, la componentes perpendicular es reflejada más eficazmente que la paralela, y para un ángulo particular, llamado ángulo de Brewster, la luz polarizada paralelamente se transmite en su totalidad. El ángulo de reflexión interna para el rayo del arco iris cae cerca del ángulo de Brewster. Por eso, la luz del arco iris presenta una polarización perpendicular.



LA DISPERSION DE ATOMOS POR ATOMOS crea un arco iris de partículas. El papel que el índice de refracción representa en la desviación óptica está figurado aquí por las fuerzas interatómicas. La principal diferencia es que estas fuerzas varían gradual y continuamente, de modo que los átomos siguen trayectorias curvadas. Cuando un átomo se acerca a otro, la fuerza que se ejerce entre ellos consiste al principio en una atracción que aumenta continuamente *(sombreado en color)* pero a muy corta distancia se hace fuertemente repulsiva *(sombreado gris)*. Un máximo local del ángulo de desviación corresponde en el caso óptimo al ángulo del arco iris. Es el ángulo bajo el cual se curva la trayectoria que mejor aprovecha la parte atractiva del potencial.



ARCO IRIS ATOMICO detectado por E. Hundhausen y H. Pauly, de la Universidad de Bonn, al estudiar la dispersión de átomos de sodio por átomos de mercurio. Las oscilaciones en el número de átomos desviados corresponden a un arco iris primario y dos máximos supernumerarios. Un arco de esta clase posee información sobre la intensidad y alcance de las fuerzas interatómicas.

de reflexión, es decir, el que contiene los rayos incidente y reflejado, suministra una referencia geométrica conveniente. Los estados de polarización de la luz incidente se pueden dividir en paralelos a ese plano y perpendiculares a él. En ambas polarizaciones, el poder reflector de la superficie es pequeño para ángulos de incidencia cercanos a la perpendicular y crece muy rápidamente en la proximidad de un ángulo crítico cuyo valor está determinado por el índice de refracción. Sobrepasado ese ángulo crítico, el rayo es totalmente reflejado con independencia de su estado de polarización. Para ángulos intermedios, sin embargo, el poder reflector depende de la polarización. A medida que la incidencia se hace más próxima a la rasante, se refleja una proporción que crece de una forma constante de la componente polarizada perpendicularmente. Para la componente paralela, por otra parte, el poder reflector decrece antes de empezar a aumentar. Cuando se trata de un ángulo en particular, el poder reflector para la onda polarizada paralelamente se anula por completo; esa onda se transmite en su totalidad. Por tanto, para luz solar incidente a ese ángulo, el rayo internamente reflejado se halla polarizado perpendicularmente al plano de reflexión. El ángulo en cuestión se denomina ángulo de Brewster, en honor de David Brewster, que explicó su importancia en 1815.

La luz del arco iris está casi completamente polarizada, como se puede ver mirando el arco iris a través de gafas de sol Polaroid y girando los cristales alrededor de la línea de visión. La fuerte polarización proviene de una notable coincidencia: el ángulo de incidencia interna para el rayo del arco iris es muy próximo al ángulo de Brewster. La mayor parte de la componente paralela escapa en los rayos de clase 2 transmitidos, dejando en el arco iris un predominio de rayos polarizados perpendicularmente.

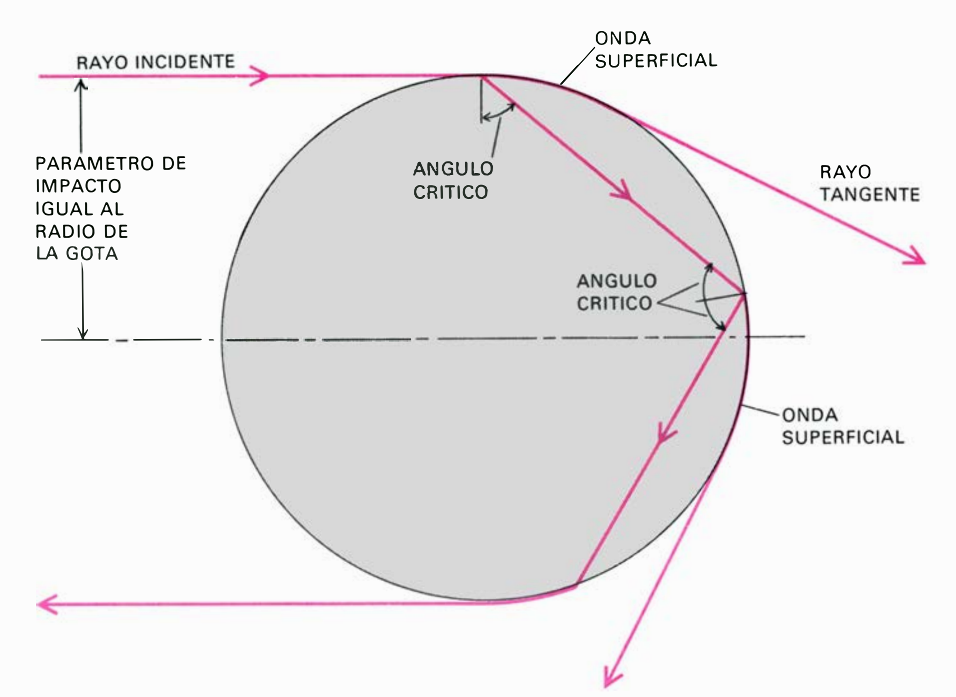
Con el conocimiento de que tanto la materia como la radiación se pueden comportar como ondas, el campo de aplicación de la teoría del arco iris se ha ensanchado. Debe abarcar ahora nuevos e invisibles arcos iris producidos en la dispersión atómica y nuclear.

Una cierta analogía entre la óptica geométrica y la mecánica clásica de partículas había sido ya percibida en 1831 por el matemático irlandés Guillermo Rowan Hamilton. Análogas a los rayos de la óptica geométrica son las trayectorias de las partículas; el cambio de dirección de un rayo de luz al penetrar en un medio de distinto índice de refracción

corresponde a la desviación de una partícula que se mueve bajo la acción de una fuerza. En la dispersión de las partículas se presentan analogías respecto de muchos efectos ópticos, entre los cuales hay que contar el arco iris.

Consideremos el choque entre dos átomos de un gas. Al acercarse los átomos partiendo de una separación inicial grande, quedan al principio sometidos a una atracción constantemente creciente. A distancias pequeñas, sin embargo, las cortezas electrónicas de los átomos comienzan a interpenetrarse y la fuerza atractiva disminuye. A distancias muy pequeñas se convierte en una repulsión cada vez más fuerte.

Como en el experimento óptico, la dispersión atómica se puede analizar trazando las trayectorias de los rayos en función del parámetro de impacto. Debido a la variación gradual y continua de las fuerzas, los átomos siguen trayectorias curvadas en lugar de cambiar bruscamente de dirección, como hacían los rayos en la superficie de separación de dos medios de distinto índice de refracción. Aun cuando algunas de las trayectorias son bastante complicadas, cada parámetro de impacto corresponde a un ángulo de desviación único; además, hay una trayectoria que representa una desviación angular local máxima. Esa trayectoria resulta ser la que hace un uso más eficaz de la interacción atractiva entre átomos. Debe esperarse una fuerte concentración de partículas en la vecindad de este ángulo: es el ángulo del arco iris para átomos que interaccionan.



LA TEORIA DEL MOMENTO cinético complejo del arco iris comienza con la observación de que un fotón, o cuanto de luz, incidente sobre una gota a cierto parámetro de impacto (que no se puede definir exactamente), transporta momento cinético. En la teoría, las componentes de ese momento cinético se extienden hasta abarcar valores complejos, es decir, valores que contienen la raíz cuadrada de −1. Las consecuencias de este método se pueden ilustrar mediante el ejemplo de un rayo que incide tangencialmente sobre una gota. El rayo genera ondas superficiales que viajan sobre la superficie de la gota y emiten radiación continuamente. El rayo puede también penetrar en la gota en el ángulo crítico para la reflexión total interna, emergiendo bien para formar otra onda superficial o bien para repetir el trayecto interno que, según se explica en el texto, es más corto.

En 1959, Kenneth W. Ford, de la Universidad Brandeis y John A. Wheeler, de la Universidad de Princeton, formularon un tratamiento mecanocuántico del arco iris atómico y nuclear. La interferencia entre trayectorias emergentes en la misma dirección da lugar a máximos supernumerarios de intensidad. Se ha deducido una teoría para la dispersión de partículas, análoga a la de Airy.

El primer arco iris atómico fue observado en 1964 por E. Hundhausen y H. Pauly, de la Universidad de Bonn, en la dispersión de átomos de sodio por átomos de mercurio. Se detectaron el máximo principal del arco iris y dos supernumerarios; en experimentos más recientes se han observado oscilaciones en escala todavía más fina. Los arcos iris medidos en esos experimentos llevan consigo información sobre las fuerzas interatómicas. De la misma manera que el ángulo del arco iris óptico depende solamente del índice de refracción, el ángulo del arco iris atómico viene determinado por la intensidad de la parte atractiva de la interacción. Del mismo modo, las posiciones de los máximos supernumerarios dependen del tamaño, y suministran información sobre el alcance de la interacción. En la actualidad se han hecho observaciones de la misma clase referentes a la dispersión de núcleos atómicos.

La teoría de Airy ha tenido muchos éxitos satisfactorios, pero comporta una desazonante laguna cognoscitiva: la necesidad de adivinar la distribución de amplitudes a lo largo del frente de onda inicial que se ha escogido. Las hipótesis empleadas para estimarlas son plausibles sólo para gotas de agua bastante grandes. En este contexto, la mejor manera de expresar el tamaño es mediante un "parámetro de tamaño", definido como la relación de la circunferencia de la gota a la longitud de onda de la luz. El parámetro de tamaño varía de aproximadamente 100 para niebla o neblina a varios millares para gotas de lluvia grandes. La aproximación de Airy, es plausible solamente para gotas con un parámetro de tamaño mayor que aproximadamente 5000.

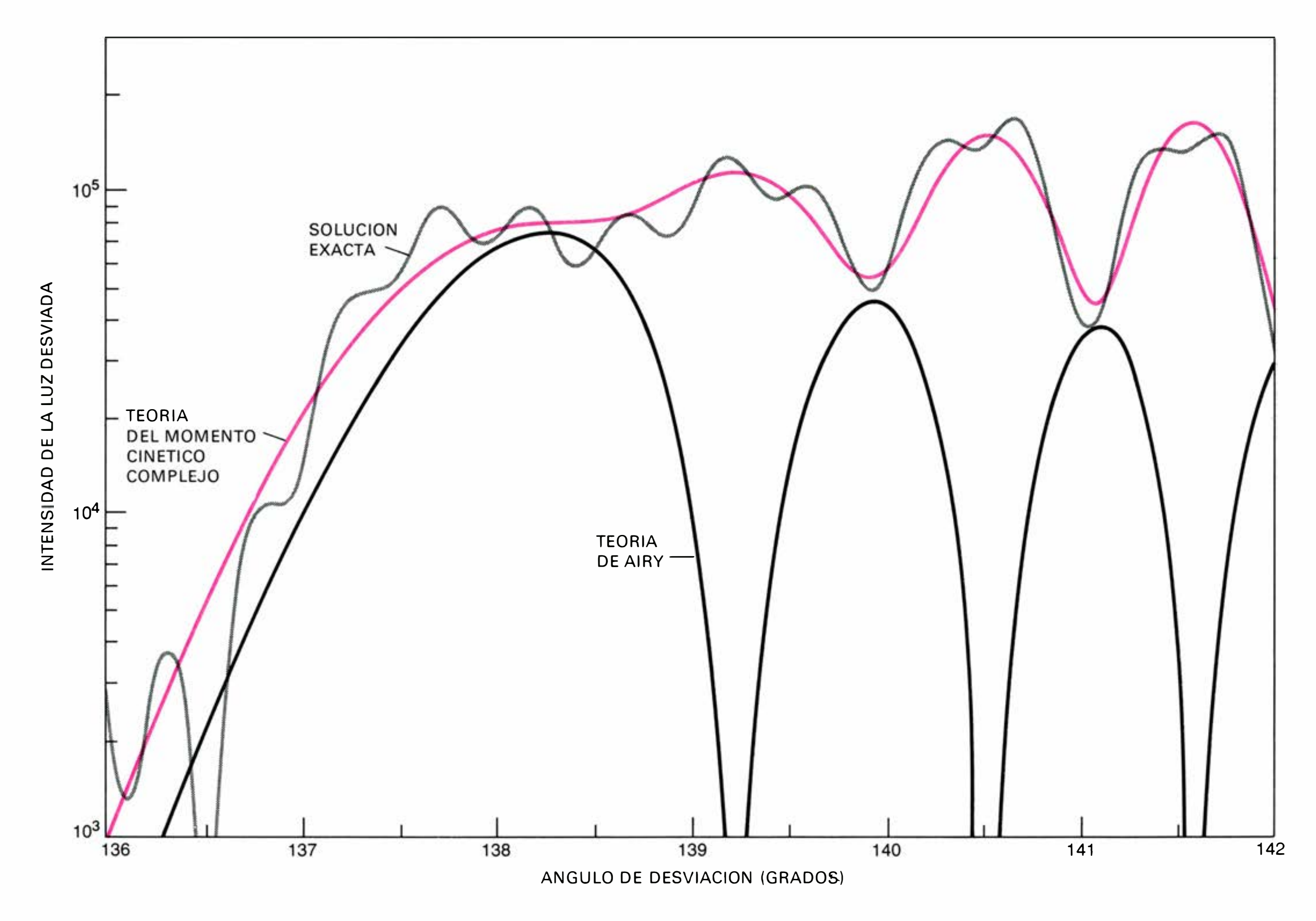
Resulta irónico que un problema tan intratable como el del arco iris tenga realmente una solución exacta, que además es conocida desde hace muchos años. En cuanto James Clerk Maxwell propuso la teoría electromagnética de la luz, hace aproximadamente un siglo, se hizo posible dar una formulación matemática precisa del problema del arco iris óptico. Se necesita calcular la dispersión de una onda electromagnética plana por una esfera homogénea. La solución de un problema similar, aunque ligeramente más sencillo, es decir, la dispersión de ondas sonoras por una esfera, había sido estudiada por varios investigadores en el siglo XIX, particularmente por Lord Rayleigh. La solución que obtuvieron constaba de una serie infinita de términos llamados ondas parciales. Una solución de la misma forma fue dada en el caso del problema electromagnético por Gustav Mie y Peter J. W. Debye en 1908.

Dada la existencia de una solución exacta del problema de la dispersión, podría parecer cosa sencilla determinar todos sus aspectos, incluyendo el carácter preciso del arco iris. El problema, naturalmente, es la necesidad de sumar la serie de ondas parciales, cada término de la cual es una función un tanto complicada. La serie se puede truncar para dar una solución aproximada, pero este método es práctico sólo en algunos ca-

sos. El número de términos que deben retenerse es del mismo orden de magnitud que el parámetro de tamaño. La serie de ondas parciales es, por tanto, muy adecuada para el tratamiento de la difusión de Rayleigh, que es la responsable del color azul del cielo; en ese caso, las partículas dispersoras son moléculas, mucho más pequeñas que la longitud de onda, de modo que basta con un término de la serie. Para el problema del arco iris, deben tomarse en cuenta parámetros de tamaño de hasta varios millares. Una buena aproximación a la solución por el método de ondas parciales requeriría evaluar la suma de varios miles de términos complicados. Se han hecho intentos con ordenadores, pero los resultados son funciones rápidamente variables del parámetro de tamaño y del ángulo de desviación, de modo que el trabajo y el costo se hacen rápidamente prohibitivos. Además, un ordenador puede solamente calcular soluciones numéricas; no ofrece medio de penetrar en la física del arco iris. Estamos, pues, en la torturante situación de conocer una forma de la solución exacta y ser, no obstante, incapaces de extraer de ella una mayor comprensión del fenómeno que describe.

Los primeros pasos hacia la resolución de esta paradoja se dieron en los primeros años del siglo XX por los matemáticos Henri Poincaré y G. N. Watson. Hallaron un método para transformar la serie de ondas parciales, que sólo converge muy lentamente hacia su límite, en una expresión rápidamente convergente. La técnica en cuestión ha llegado a ser conocida como transformación de Watson o método del momento cinético complejo.

No es particularmente difícil ver por qué el momento cinético está relacionado con el problema del arco iris, aunque es menos evidente que sea preciso considerar valores complejos del momento cinético. La explicación más sencilla viene suministrada por una teoría corpuscular de la luz, en la que un haz de luz debe considerarse como un chorro de partículas llamadas fotones. Aunque el fotón no tiene masa, transporta energía y cantidad de movimiento en proporción inversa a la longitud de onda de la correspondiente onda luminosa. Cuando un fotón choca con una gota de agua con un parámetro de impacto mayor que cero, el fotón transporta un momento cinético igual al producto de su cantidad de movimiento por el parámetro de impacto. Cuando el fotón sufre una serie de reflexiones internas, está describiendo, en realidad, elementos de órbita en torno al centro de la gota. La mecánica cuántica impone de hecho ciertas restric-



LAS TEORIAS CUANTITATIVAS del arco iris predicen la intensidad de la luz desviada en función del ángulo de desviación y también del tamaño de la gota y la polarización. Aquí se presentan las predicciones de tres teorías para el caso de la luz polarizada paralelamente y desviada por gotas de circunferencia igual a 1500 veces la longitud de onda de la luz. Una curva representa la solución "exacta" del problema del arco iris, deducida de las ecuaciones obtenidas por James Clerk Maxwell para describir la radiación electromagnética. La solución exacta es la suma de una serie infinita de términos, aproximada aquí sumando más de 1500 términos complicados para cada punto usado en el trazado de la curva. La teoría de Airy está en claro desacuerdo con la solución exacta, particularmente en la región angular de los arcos supernumerarios. Allí, la solución exacta presenta mínimos en la posición de los máximos de Airy. Por otra parte, los resultados obtenidos por el método del momento cinético complejo corresponden muy aproximadamente a la solución exacta salvo que no reproducen las oscilaciones pequeñas de alta frecuencia. Estas fluctuaciones se encuentran íntimamente asociadas con otro fenómeno óptico en la atmósfera, la gloria.

ciones adicionales sobre este proceso. Por una parte, requiere que el momento cinético tome solamente ciertos valores discretos; por otra, deniega la posibilidad de determinar en forma precisa el parámetro de impacto. Cada valor discreto del momento cinético responde a un término de la serie de ondas parciales.

Para realizar la transformación de Watson deben introducirse valores del momento cinético habitualmente considerados como imposibles físicos. Por un lado, se debe permitir al momento cinético que varíe continuamente, en lugar de hacerlo por unidades cuantificadas; y, más importante, debe permitírsele que se extienda al campo de los números complejos, es decir, los que incluyen una componente real y una imaginaria que contiene algún múltiplo de la raíz cuadrada de –1. El plano definido por esas dos componentes se designa como plano del momento cinético complejo.

Mucho es lo que se gana a cambio de las abstracciones matemáticas del método del momento cinético complejo. En particular, tras recorrer el plano del momento cinético complejo mediante la transformación de Watson, se pueden redistribuir las contribuciones a la serie de ondas parciales. En lugar de un gran número de términos, se puede trabajar con sólo unos pocos puntos llamados polos y puntos de silla en el plano del momento cinético complejo. A lo largo de los últimos años, los polos han atraído gran interés teórico en física de partículas elementales. En tal contexto se llaman generalmente polos de Regge, en honor del físico italiano Tullio Regge.

Tanto los polos como los puntos de silla tienen interpretación física en el problema del arco iris. Las contribuciones de los puntos de silla reales están asociadas a los rayos luminosos corrientes y reales que hemos estado considerando a lo largo de este artículo. ¿Qué ocurre con los puntos de silla complejos? Los números imaginarios o complejos se consideran ordinariamente como soluciones de una ecuación que no tiene sentido físico, pero no son soluciones sin sentido. En las descripciones de la propagación de ondas, las componentes imaginarias van usualmente asociadas al amortiguamiento de la amplitud de la onda. Por ejemplo, en la reflexión total interna de un rayo de luz en la superficie de separación agua-aire, una onda luminosa "pasa a través del espejo". Su amplitud, sin embargo, se amortigua rápidamente, de modo que la intensidad se hace despreciable a una profundidad del orden de una sola longitud de onda. Tal onda no

se propaga en el aire; por el contrario, queda ligada a la superficie de separación entre agua y aire, propagándose a lo largo de ella; se llama onda evanescente. La descripción matemática de la onda evanescente lleva consigo las componentes imaginarias de una solución. El efecto túnel de mecánica cuántica, en el cual una partícula pasa a través de una barrera de potencial sin superarla, tiene una base matemática análoga. También aparecen "rayos complejos" en el lado de sombra de una cáustica, donde describen la amplitud amortiguada de las ondas luminosas difractadas.

Las contribuciones de los polos de Regge a la serie de ondas parciales transformada van asociadas a ondas superficiales de otra clase. Estas ondas son excitadas por rayos incidentes que atacan la esfera tangencialmente. Una vez creada tal onda, se propaga sobre la esfera, pero se va amortiguando continuamente porque emite radiación en sentido tangencial, como un aspersor de jardín. En cada punto, a lo largo del camino circular de la onda, ésta penetra también en la esfera con un ángulo de incidencia igual al crítico para reflexión total interna, reemergiendo como onda superficial después de describir uno o más de tales atajos. Es interesante observar que Johannes Kepler conjeturó, en 1584, que esta clase de rayos en "molinillo de viento" podrían ser responsables del arco iris, pero abandonó la idea porque no le conducía al ángulo correcto del arco.

En 1937, los físicos holandeses Balthus van der Pol y H. Bremmer aplicaron la transformación de Watson al problema del arco iris, pero sólo pudieron demostrar que se podía obtener la aproximación de Airy como caso límite. En 1965, el autor desarrolló una versión mejorada del método de Watson y la aplicó al problema del arco iris en 1969 con algo más de éxito.

En el sencillo análisis cartesiano vimos que en el lado iluminado del arco iris hay dos rayos que emergen en la misma dirección; para el ángulo del arco iris, se funden en el rayo único de mínima desviación de Descartes, y en el lado en sombra desaparecen. En el plano del momento cinético complejo, como he mencionado, cada rayo geométrico corresponde a un punto de silla real. De aquí que, en términos matemáticos, un arco iris es meramente la intersección de dos puntos de silla en el plano del momento cinético complejo. En la región de sombra más allá del ángulo del arco iris, no es que los puntos de silla simplemente desaparezcan, sino que se hacen complejos, es decir, crean parte imaginaria. La luz difractada en la banda oscura de Alejandro proviene de un punto de silla complejo. Es un ejemplo de "rayo complejo" en el lado en sombra de una curva cáustica.

Debe observarse que la adopción del método del momento cinético complejo no implica que las anteriores soluciones del problema del arco iris fueran erróneas. La explicación de Descartes del arco primario como rayo de mínima desviación no es en manera alguna incorrecta, y los arcos supernumerarios siguen pudiendo considerarse como un efecto interferencial, tal como Young había propuesto. El método del momento cinético complejo simplemente da cuenta más completa de los caminos permitidos a un fotón en la región del cielo correspondiente al arco iris, y por ello logra resultados más reales.

En 1975, Vijay Khare, de la Universidad de Rochester, hizo una comparación detallada de las tres teorías del arco iris: la aproximación de Airy, la solución "exacta" obtenida por sumación de la serie de ondas parciales mediante ordenador, y los términos del arco iris en el método del plano del momento cinético complejo, asociados con la intersección de dos puntos de silla. Para la polarización perpendicular dominante, la teoría de Airy requiere solamente pequeñas correcciones en el arco primario, y sus errores se hacen apreciables solamente en la región de los arcos supernumerarios. Para los rayos desviados que están polarizados perpendicularmente al plano de incidencia, sin embargo, la aproximación de Airy falla claramente. Para los arcos supernumerarios, la solución exacta muestra mínimos donde la teoría de Airy presenta máxima intensidad y viceversa. Este serio fallo es un resultado de la casi coincidencia entre el ángulo de reflexión interna para los rayos del arco iris y el ángulo de Brewster. Para este ángulo, la amplitud del rayo refractado cambia de signo, cambio que la teoría de Airy no tiene en cuenta. Como resultado del cambio de signo, la interferencia a lo largo de las direcciones que corresponden a los máximos en las soluciones de Airy es destructiva en lugar de constructiva.

En términos de las características en gran escala, tales como el arco primario, los arcos supernumerarios y el aspecto de la difracción en el lado oscuro, el resultado a partir del momento cinético complejo concuerda muy bien con la solución exacta. Las fluctuaciones de intensidad a menor escala en la curva exacta no son tan bien reproducidas por los términos de arco iris en el método del momento cinético complejo. Por otra parte, la solución exacta, para un parámetro de tamaño típico de 1500, requiere la sumación de más de 1500 términos complicados; la curva del momento cinético complejo se obtiene de sólo unos pocos términos mucho más sencillos.

Las pequeñas fluctuaciones residuales en la curva exacta de intensidad provienen de reflexiones internas de orden superior, es decir, de rayos pertenecientes a las clases 3 y 4. Estos tienen poco interés para el arco primario, mas para mayores ángulos de desviación su contribución aumenta, y en la proximidad de la dirección de retrodifusión se hace dominante. Estos rayos son los responsables de otra fascinante manifestación meteorológica: las "glorias".

La gloria aparece como un halo de colores espectrales que rodea la sombra que un observador proyecta sobre nubes o niebla; cuando más comúnmente se observa es desde un avión que vuela por encima de nubes. También se puede explicar mediante la teoría del momento cinético complejo, pero la explicación es más complicada que la del arco iris. Un grupo de contribuciones a la gloria proviene de las ondas superficiales descritas por los polos de Regge que van asociadas a los rayos tangenciales del tipo de molinillo de viento de Kepler. Las reflexiones internas múltiples que dan lugar a polígonos estrellados cerrados desempeñan un importante papel, dando lugar a resonancias o refuerzos de la intensidad. Tales coincidencias geométricas encajan muy bien en el espíritu de las teorías de Kepler.

Un segundo e importante grupo de contribuciones, puesto de manifiesto por Khare, viene del lado sombreado de arcos iris de orden superior que aparecen cerca de la dirección de retrodifusión. Esas contribuciones representan el efecto de rayos complejos. El arco iris de décimo orden, formado a sólo unos pocos grados de distancias angular de la dirección de retrodifusión, es particularmente efectivo.

Para los arcos iris de orden superior, la teoría de Airy daría resultados incorrectos para ambas polarizaciones, y por tanto debe utilizarse la teoría del momento cinético complejo. Se podría así decir que la gloria está en parte formada por la sombra de un arco iris. Satisface descubrir en la elegante, pero aparentemente abstracta teoría del momento cinético complejo, una explicación para esos dos fenómenos naturales, y encontrar además una inesperada relación.

Juegos matemáticos

El triángulo de bolas, una paradoja versificada y un surtido de problemas

Martin Gardner

Presento aquí un revoltillo de ocho problemas, acompañados de una nota acerca del popular juego de tablero llamado Otelo y de la discusión del problema del mes pasado. Las soluciones de estos ocho problemas se darán en el próximo número.

1. El triángulo de bolas de billar americano. Me cuenta el coronel George Sicherman, de Buffalo, que hace unos 10 años, al observar una partida de billar, se le ocurrió el siguiente problema: ¿Es posible formar un "triángulo de diferencias" al situar las bolas en la disposición triangular habitual antes de comenzar una partida? En un triángulo de diferencias se sitúan los números de 1 a 15 de modo que, debajo de cada par, de números se halle su diferencia positiva.

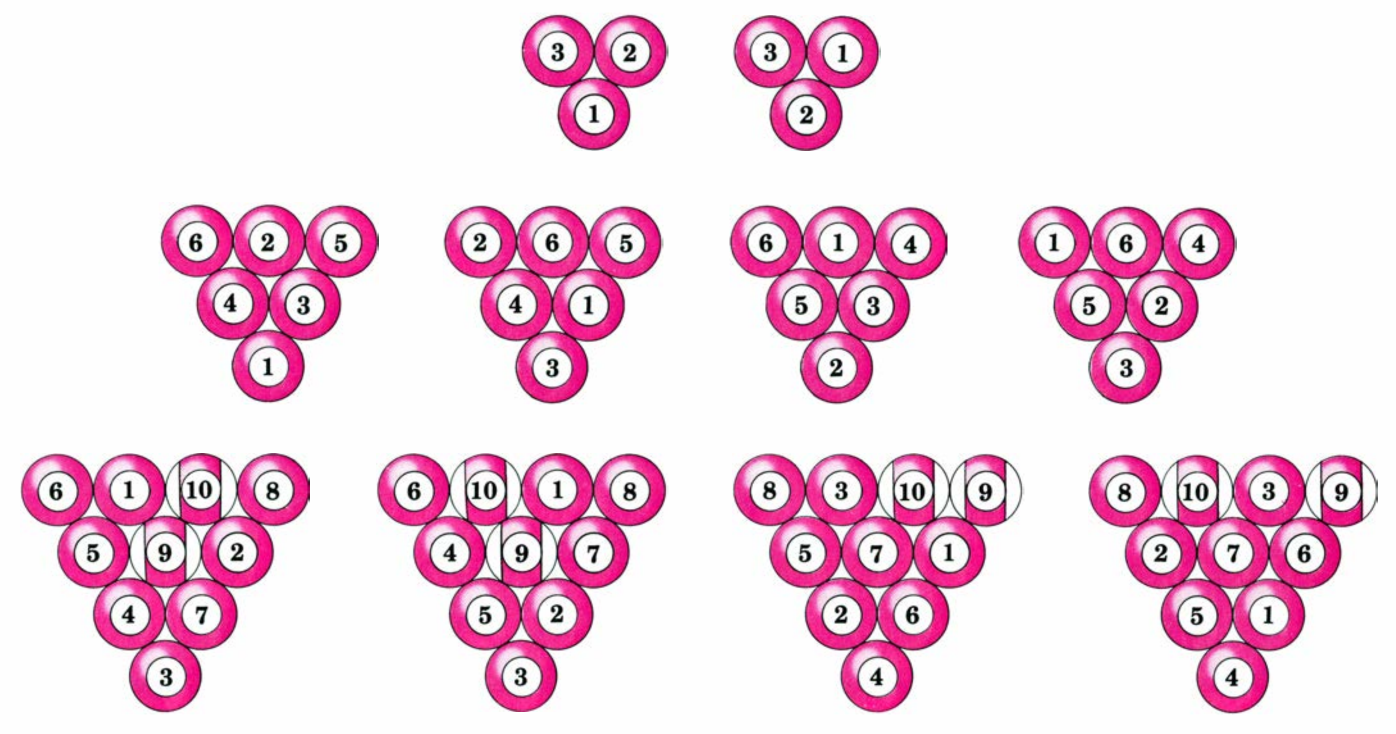
Es evidente, por la ilustración inferior, que el problema es trivial con las bolas numeradas, 1, 2 y 3, y que con ellas se pueden obtener dos soluciones. La ilustración muestra también las cuatro soluciones posibles con seis bolas, y las cuatro correspondientes a 10. Para sorpresa de Sicherman, con las 15 bolas se tiene solamente una solución fundamental (salvo simetría). ¿Podría hallarla el lector?

La búsqueda de la solución se simplifica notablemente estudiando en primer lugar disposiciones triangulares de pares e impares, a fin de ver en cuáles de ellas hay precisamente ocho lugares impares y siete pares. No se tarda mucho en descubrir que solamente hay cinco posibles disposiciones de la hilera superior del triángulo: PPIPI, IPPPI, IIIPP, IIIIP, IIPPI. Es evidente que la bola número 15 ha de hallarse en la hilera superior y que la bola número 14 ha de estar, o bien en la misma fila, o bien debajo de la 15 y la 1. Otras tretas como ésta abrevian un análisis agotador.

Este problema se relaciona con otro expresado por Hugo Steinhaus en su libro *One Hundred Problems in Elementary Mathematics.* Dada una disposición triangular con un número par de puntos, ¿será siempre posible formar una configuración de diferencias par-impar en la que el número de puntos pares iguale al de impares? El problema permaneció sin resolver durante más de una década hasta que Heiko Harborth, en *Journal of Combinatorial Theory (A),* vol. 12, 1972, pp. 253-259, probó que la repuesta es afirmativa.

Que yo sepa no se ha hecho ninguna labor sobre lo que podríamos llamar el problema general de las bolas del billar americano. Con cualquier número triangular de bolas, consecutivamente numeradas a partir de 1, ¿será siempre posible formar un triángulo de diferencias? Y si no, ¿cuál sería el mayor triángulo para el que existe solución? Si lo hay, ¿cuál es? Ya sabemos que existen configuraciones par-impar, para dichas soluciones, para todos los triángulos con un número par de bolas. ¿Existen también para todo triángulo con un número impar de bolas?

Añadamos el siguiente problema para los lectores que tengan éxito con el de las 15 bolas. Supongamos que las bolas

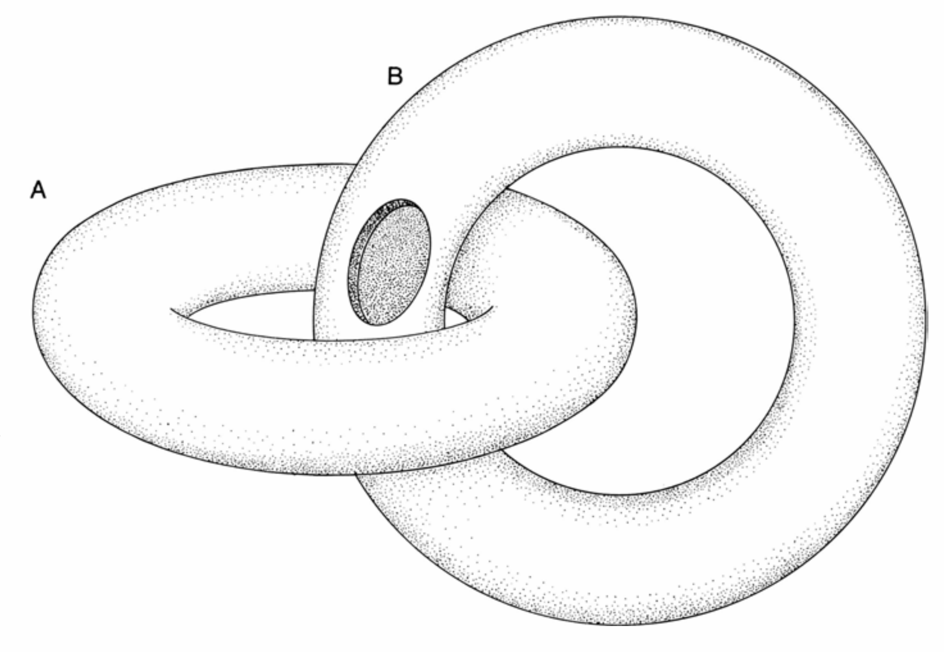


Triángulos de diferencias de tres, seis y diez bolas

lleven los 15 números pares consecutivos que van del 2 al 30 inclusive. ¿Se podría disponer el conjunto en un triángulo de diferencias?

2. Canibalismo toroidal. Toro es una superficie en forma de rosquilla. Imaginemos un neumático. Es de sobra conocido que, si tuviera un agujero, podría ser vuelto del revés a través del agujero.

John Stillwell, matemático de la Universidad de Monash, en Australia, planteó el problema siguiente. Dos toros, A y B, están unidos como se muestra en el dibujo superior de esta página. Hay una "boca" (un agujero) en B. Podremos estirar, comprimir y deformar cualquiera de ellos tanto como queramos, pero, por supuesto, sin romperlos. ¿Podría B tragarse al A? Al final B ha de tener su forma primitiva, aunque será mayor, y A deberá estar completamente metido en B.



¿Podrá el toro B tragarse al A?

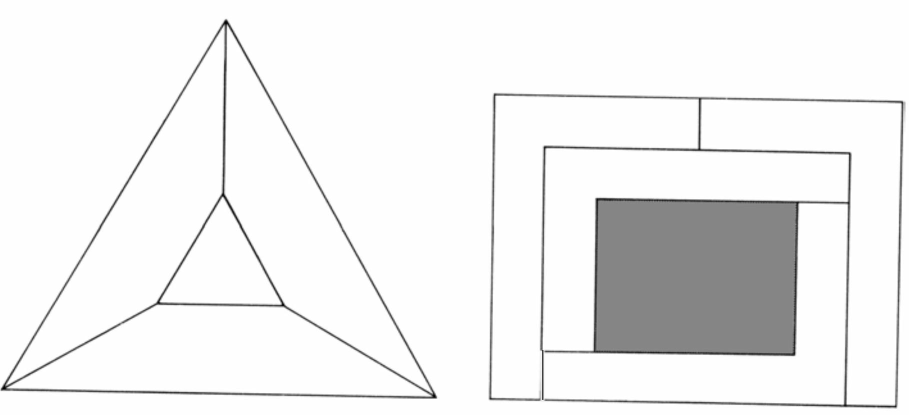
3. Explorando tétradas. La noticia más sensacional del año pasado en matemática recreativa fue seguramente el anuncio hecho por dos matemáticos de la Universidad de Illinois de haber demostrado la conjetura del mapa de 4 colores. Esta famosa conjetura se confunde frecuentemente con un teorema topológico más simple, que se prueba fácilmente; éste anuncia que no más de 4 regiones del plano pueden tener frontera común. Michael R. W. Buckley, en *Journal of Recreational Mathematics,* vol. 8 (1975), propuso el nombre de tétrada para cuatro regiones planares simplemente conexas, que, dos a dos, comparten una porción finita de una frontera común.

La tétrada de la izquierda en la ilustración central de esta página no tiene huecos. Nótese que solamente 3 regiones son congruentes. La tétrada de la derecha tiene 4 regiones congruentes y un hueco grande. ¿Es posible, preguntó Buckley, construir una tétrada con cuatro regiones congruentes y sin agujero?

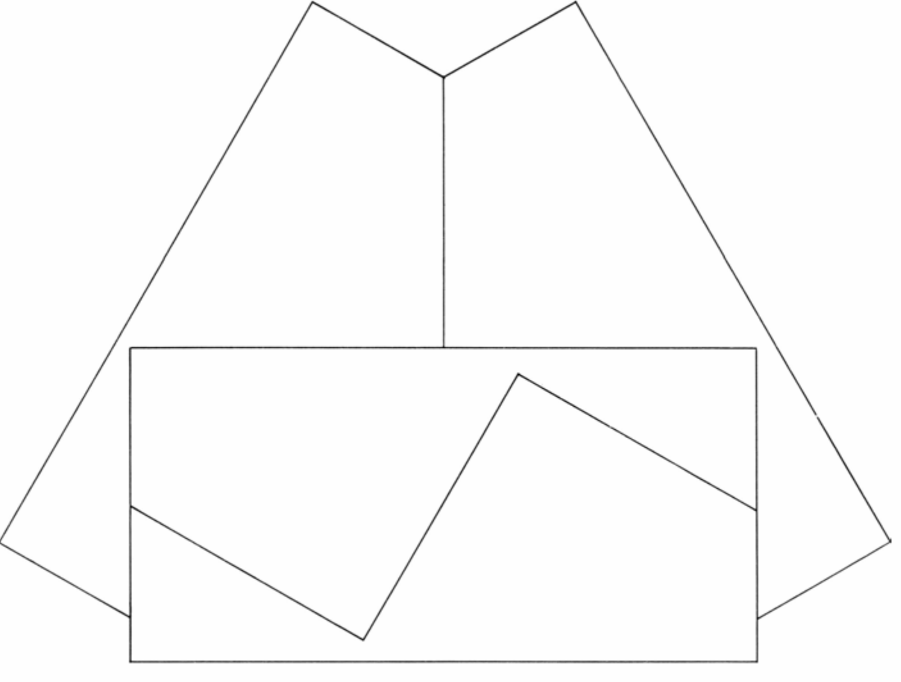
Esta pregunta ha sido contestada afirmativamente por Scott Kim, estudiante de la Universidad de Stanford. No se han publicado sus resultados y le agradezco que me permitiera facilitar aquí algunos de ellos.

La ilustración inferior de esta página muestra una solución con cuatro hexágonos congruentes. Se desconoce si se puede obtener solución con un polígono de menos de 6 lados o si hay alguna solución cuya línea exterior sea convexa.

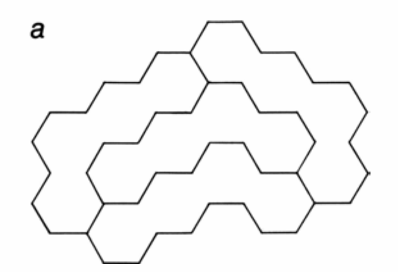
La parte *a* de la figura de la página 98 muestra una solución con polihexos congruentes de orden 4. (Un polihexo es una unión de hexágonos regulares coin-

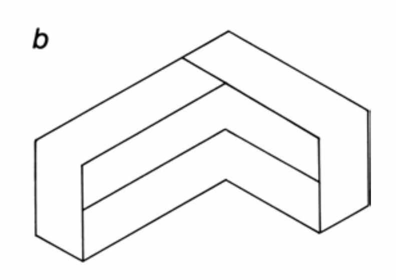


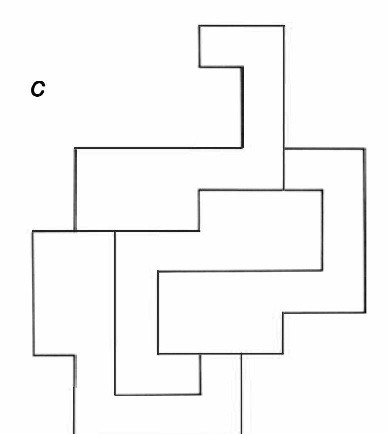
Un par de tétradas

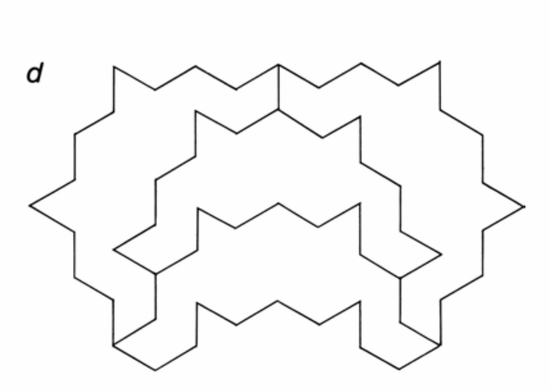


Una tétrada formada por hexágonos congruentes









Solución de distintas tétradas

cidentes.) Es fácil demostrar que no hay solución posible con polihexos de orden inferior.

La parte *b* de la ilustración muestra una solución con poliamones congruentes de orden 10. (Un poliamón es una unión de triángulos equiláteros congruentes.) No se sabe si hay solución con poliamones de orden inferior.

La parte *c* de la ilustración muestra una solución con poliominos congruentes de orden 12. (Un poliomino es una unión de cuadrados congruentes.) Tampoco se sabe si hay solución con poliominos de orden inferior.

La parte *d* muestra una solución con piezas congruentes con simetría axial y frontera igualmente simétrica. ¿Existe tal solución con polígonos de menos lados?

Me complacería tener noticia de cualquier progreso que los lectores sean capaces de lograr con respecto a los resultados anteriormente expuestos.

Ya se sabía desde alrededor de 1870 que, en 3 dimensiones, puede acoplarse un número infinito de sólidos congruentes de modo que cada par comparta una parte común de superficie. Para lectores no familiarizados con este resultado, he aquí un problema cuya solución se dará el mes próximo. Mostrar de qué modo un número infinito de policubos congruentes pueden acoplarse sin dejar huecos interiores, de modo que dos a dos compartan una porción de superficie. (Un policubo es una unión de cubos congruentes.) Una tal estructura demuestra que se requiere un número ilimitado de colores para colorear cualquier "mapa" de 3 dimensiones.

4. Buenos y malos. Raymond Smullyan, matemático de la City University of New York, cuyos notables problemas de ajedrez ya son familiares a los lectores habituales de este espacio, es el responsable de estos cuatro atractivos acertijos lógicos con buenos y malos, y tal vez algunas personas más. En todos ellos, el bueno siempre dice la verdad y el malo siempre miente.

A dice: "B es bueno".
B dice: "A no es bueno."
Pruébese que uno dice la verdad, pero no es bueno.
A dice: "B es bueno"
B dice: "A es malo".
Pruébese que, o bien uno de ellos dice la verdad pero no es bueno, o bien uno miente pero no es malo.

En los problemas arriba mencionados habrá que considerar la posibilidad de un tercero que no sea ni bueno ni malo. En los dos problemas siguientes, cada uno de los tres personajes implicados es, o bueno, o malo.

C dice: "B es malo".
B dice: "A y C son del mismo tipo (ambos buenos o ambos malos)".
¿Qué es A?

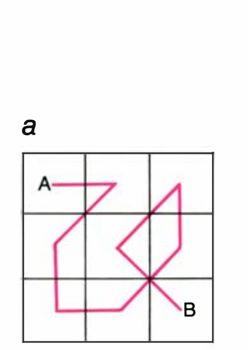
A dice: "B y C son del mismo tipo".
Se le pregunta a C: "¿Son A y B del mismo tipo?"
¿Qué responde C?

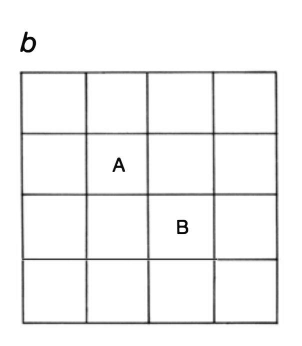
5. Los recorridos del rey perdido. Hace varios años, Scott Kim propuso lo que él denomina "el recorrido del rey perdido". Se trata del recorrido de un rey sobre un pequeño tablero de ajedrez y sujeto a las siguientes condiciones:

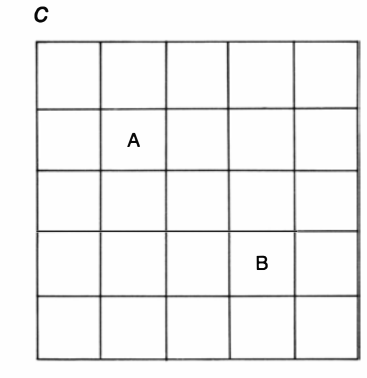
Primero, el rey debe pasar por cada celdilla una y sólo una vez.

Segundo, el rey debe cambiar de dirección después de cada movimiento, esto es, que no puede moverse dos veces consecutivas en la misma dirección.

Tercero, el número de puntos por donde la trayectoria del rey se corta a sí misma debe reducirse al mínimo.







Recorridos del rey en tres tableros de distinto tamaño

La parte *a* de la ilustración inferior de la página 98 muestra el único recorrido posible en un tablero de 3×3 desde la casilla A a la B. Tiene un cruce y es único, salvo, evidentemente, simetría respecto a la diagonal principal. Un recorrido cerrado es imposible sobre este tablero. Se encuentran fácilmente recorridos cerrados y sin cruces en tableros de 4×4. En los de 5×5 un recorrido cerrado probablemente exija dos cruces. El problema es menos interesante en dameros mayores, puesto que se cree que siempre son posibles recorridos cerrados o abiertos entre dos casillas cualesquiera, sin cruces.

Veamos dos bonitos problemas de recorridos planteados por Kim:

Sobre el tablero de 4×4, que se muestra en la parte *b* de la ilustración, encuéntrese un recorrido de rey perdido de A a B con sólo 3 cruzamientos. La solución es única.

En el tablero de 5×5 de la parte *c* del dibujo encuéntrese un recorrido de rey perdido de A a B con sólo 2 cruces. Este problema es de una dificultad insólita. Kim no sabe si la solución que él encontró es única o si el problema se puede resolver con un solo cruce.

6. Elipses de Steiner. Esta reliquia se remonta a Jakob Steiner, notable geómetra suizo del siglo XIX. Me baso, para recordarlo, en que es uno de los mejores ejemplos que conozco de problema de difícil resolución analítica, pero que es rabiosamente fácil enfocándolo con un cierto retorcimiento mental y algún conocimiento de geometría elemental, plana y proyectiva.

Se da un triángulo de lados 3, 4 y 5. Su área es de 6 unidades cuadradas. Se desea calcular cuál sería el área mínima de la elipse circunscrita a él y el área máxima de la elipse inscrita en él.

7. Distancias diferentes. Es fácil colocar 3 fichas en las casillas de un damero de 3×3 de modo que no haya nunca 2 pares de fichas a la misma distancia. Se supone que cada ficha señala el centro exacto de una casilla y que las distancias se miden sobre una línea recta que une los centros. Salvo giros y simetrías, existen tres soluciones que figuran en la ilustración inferior de la página opuesta, como puede verse.

También es fácil poner 4 fichas en un tablero de 4×4 de modo que todas las distancias entre pares sean diferentes. Hay 16 formas de hacerlo. Sobre un tablero de 5×5 la cifra se eleva a 28.

En el *Journal of Recreational Mathematics,* de enero de 1972, Sidney Kravitz pidió solución para cuadrados de orden 5 y 6. La solución para el cuadrado de orden 7 resultó difícil porque por primera vez entra en escena el triángulo rectángulo de lados 3, 4 y 5 (la terna pitagórica mínima). El número de posibles disposiciones queda muy reducido debido a que son posibles distancias de 5 unidades tanto en filas y columnas como en diagonal. Como lo descubrieron unos lectores de Scientific American, sólo hay dos soluciones. Aparecen en la parte de abajo de la ilustración inferior de esta página.

¿En qué cuadrados de lado *n* es posible colocar *n* fichas de modo que todas las distancias sean distintas? Como se informaba el año pasado en el número de otoño del *Journal* antes citado, John H. Muson demostró que no había soluciones posibles en tableros de orden 16 o superior. Harry L. Nelson, nuevo editor de la revista, bajó el límite a 15. Milton W. Green, con un programa de búsqueda exhaustiva mediante computador, estableció la imposibilidad para los de orden 8 y 9 y encontró una solución única para el de orden 7. David Babcock, un editor de *Popular Computing,* confirmó estos resultados hasta orden 8. El problema quedó liquidado en noviembre pasado cuando Michael Beeler, con su programa de computador, confirmó los resultados anteriores y demostró su imposibilidad para los de orden 10 a 14.

El tablero de orden 7, así pues, es el mayor para el que existe solución. La solución es única y muy difícil de hallar sin computador. No obstante, los lectores podrían divertirse buscándola.

El *Journal of Recreational Mathematics,* ya en su noveno año, sigue repleto de nueva y estimulante temática. (Precios de suscripción en EE.UU. por los cuatro números anuales, $10 para suscriptores particulares y $21 para instituciones; las suscripciones pueden conseguirse de Baywood Publishing Co., 43 Central Drive, Farmigdale, N.Y. 11735.)

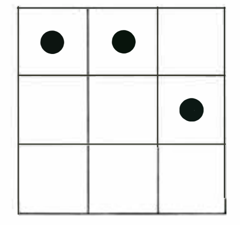
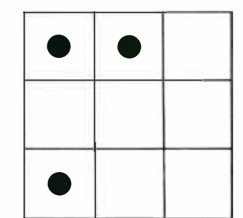
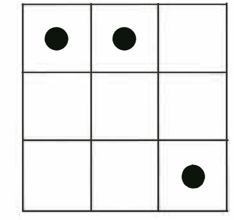
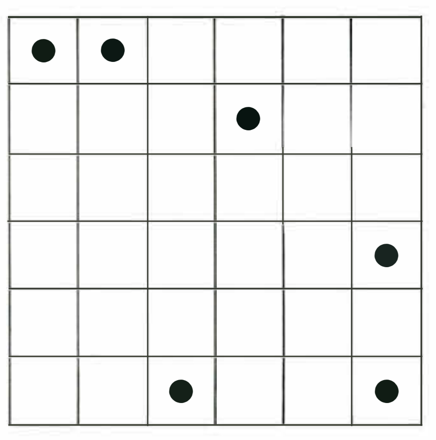
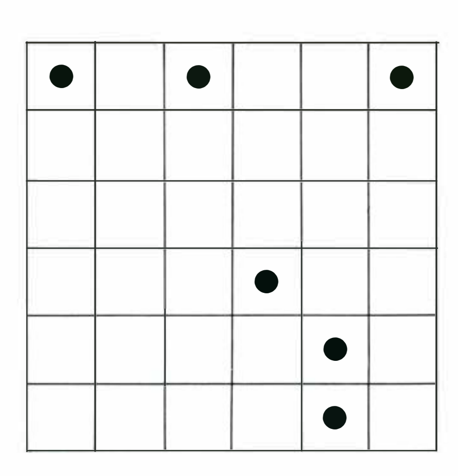
8. La paradoja versificada. Una variante divertida de la vieja paradoja del mentiroso apareció hace unos años en la revista mensual inglesa *Games and Puzzles.* Se presenta aquí como la última de 4 "aleluyas".

En el Japón una joven había
cuyas estrofas medir no se podía.
Cuando alguien el porqué preguntaba,
ella para responder suspiraba,
"Es porque yo siempre trato de
meter tantas palabras en la línea
última como posible me sería".

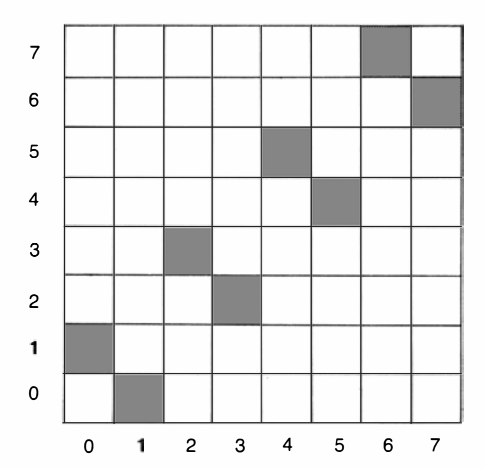
Otro joven poeta de la China
tenía para el ritmo un sentido cosa fina.
Su verso se dirigía
a un punto que concluía
repentinamente.

Había una señorita muy rotunda
cuyo verso se detenía en la segunda.

Había un joven de Verdún.

Problema de colocación



Casillas seguras del nim torre-caballo

La popularidad del "nuevo" juego de tablero Otelo motiva esta nota. Según *Time* (22 de noviembre de 1976, página 97), el juego lo inventó en 1971 un vendedor japonés, y fue llamado Otelo por su padre, erudito en Shakespeare, porque mostraba "alternativas dramáticas". Después de que el juego hubiera hecho furor en Japón en 1973 (se vendieron millones de juegos), Gabriel Industries adquirió los derechos para Estados Unidos y ahora es de los más vendidos en ese país.

Otelo no es sino un juego inglés del siglo XIX, llamado *reversi*, con el nombre cambiado. Todavía se vende en Inglaterra bajo su nombre primitivo. Los lectores que quieran evitar el coste de un pequeño tablero de plástico con sus correspondientes fichas de póker, podrán jugarlo en cualquier tablero de ajedrez o damero, proveyéndose de fichas de dos colores.

Como escribí en 1960, el *reversi* es un "juego que combina una estructura compleja con unas reglas de simplicidad maravillosa, y un juego que no merece el olvido". N. J. D. Jacobs, matemático del grupo de proceso de datos del CERN (Centro Europeo de Investigación Nuclear, 1211 Ginebra 23, Suiza), tiene un excelente programa de computador para el *reversi* y le gustaría tener noticias de cualquiera que estuviera dispuesto a jugar contra él, aportando o no su propio programa.

La tarea del mes pasado consistía en analizar un juego (similar al nim) en el que los jugadores pueden coger fichas de cualquiera de dos montones o coger una ficha de uno y dos del otro. Gana el último en jugar. En el modelo de tablero de ajedrez ilimitado explicado el mes pasado respecto del nim de W. A Wythoff, la primera regla equivale al movimiento de una torre hacia el oeste o el sur y la segunda equivale al salto de un caballo hacia el suroeste. El juego de retirar fichas es, por tanto, isomorfo al juego de acorralar una pieza de ajedrez que combinara la capacidad de la torre y el caballo.

Si la pieza se moviera sólo como una torre, el juego sobre un tablero de ajedrez sería igual al nim normal de dos montones. Pares seguros son dos enteros positivos iguales cualesquiera. Estos corresponden a las casillas de la diagonal principal del tablero que pasa por las casillas de las esquinas 0/0 y 7/7. A partir de allí su estrategia es moverse siempre hacia la diagonal. En el juego de fichas esto se traduce en mantener los montones iguales. Los pares seguros son simplemente 1/1, 2/2, 3/3...

Sorprendentemente, el darle a la torre la capacidad adicional del caballo no modifica la estrategia. Aplicando la técnica recursiva explicada el mes pasado, vemos que las casillas seguras (o pares seguros) son exactamente los mismos que en el juego de la torre.

La forma negativa del nim torre-caballo (el último en jugar pierde) es más interesante. Los pares seguros son 0/1, 2/3, 4/5, 6/7... En un tablero de ajedrez, éstos son las casillas grises de la figura de la izquierda. El que sale lleva las de ganar, pero debe poner la torre-caballo en una casilla adyacente a la del ángulo superior derecho. A partir de allí ha de mover para ocupar una casilla segura. Este procedimiento lo lleva finalmente a 0/1 o 1/0, obligando a su contrincante a realizar el movimiento final.

Los lectores podrían disfrutar analizando el juego sobre un tablero normal de ajedrez cuando la pieza colocada tiene otras capacidades del juego de ajedrez, limitando en cada caso los movimientos al oeste, sur y suroeste. Con una superdama (combinación de reina y caballo) se tiene una derrota para el que sale, tanto en el juego normal como en el inverso. Un rey lleva las de perder para el que sale en el juego normal, pero gana en el inverso. Se llega al mismo resultado si la pieza es un rey-caballo o un rey-torre. El que sale gana en ambos tipos de juego si la pieza es un rey-alfil.

Un alfil o alfil-caballo da poco juego, porque no puede moverse desde ninguna casilla de la fila anterior, ni desde la columna más alejada de la izquierda. Cualquiera de ambas piezas obviamente daría tablas en ambos tipos de juego si la finalidad es el conseguir 0/0. Si el ganar consiste en ser el último en jugar, el que sale gana en el juego normal colocando cualquier pieza en 0/7 o 7/0; y gana en el juego inverso colocando cualquier pieza en 1/7 o 7/1.

Taller y laboratorio

Los placeres y los problemas de la cría de caracoles en casa

C. L. Stong

El verano pasado puse un centímetro de barro procedente de un estanque en un frasco de vidrio, añadí agua del estanque y dejé el frasco en el alféizar de una ventana. Muy pronto el agua comenzó a rebullir de microorganismos, estando al parecer cada especie empeñada en cargar el equilibrio ecológico en su favor. Luego apareció un par de manchitas como cabezas de alfiler. Se adherían a la pared del frasco, y, al moverse, dejaban huellas transparentes sobre el vidrio empañado. En un plazo notablemente breve, las manchitas acabaron con mi diversión; al madurar, se habían transformado en un par de aburridos y voraces caracoles que, aparentemente, se comieron o envenenaron a todos los demás organismos.

Me quejé de los caracoles a Robert A. Brown, que había pertenecido a la New York Zoological Society y dirige ahora la Anti-Cruelty Society. "¡Pero si los caracoles son unos animales estupendos!" exclamó. "No los minusvalores hasta no conocer a uno personalmente." Luego me redactó un procedimiento para llegar a conocer los caracoles, que es, en esencia, el que transcribo.

Los caracoles se comportan de una manera curiosa y estrafalaria. Por ejemplo, se sumergen en un estado de latencia al encontrarse con distintas y numerosas condiciones ambientales desfavorables, tales como un exceso de calor, de frío o de sequedad. Los caracoles terrestres son especialmente fáciles de observar y de cuidar. Representan a uno de los grupos zoológicos más importantes, los moluscos. Se han desarrollado unas 80.000 especies de moluscos, de las cuales sobrevive la mitad.

De los varios miles de tipos de caracol terrestre, quizá el más fácil de mantener y observar en mi país, Norteamérica, sea el *Otala lactea.* Se envía en cestas de mimbre, como alimento, a Nueva York y otros centros metropolitanos, desde el norte de Africa, principalmente desde Marruecos. Es un pariente próximo del "escargot", el apreciado caracol comestible de los gourmets franceses. Se dice que París consume más de 200 millones de escargots por año.

O. lactea es algo más pequeño que el "escargot"; el interior de la abertura de la concha es de color castaño oscuro. Si no se encuentra esta especie en particular, sépase que el procedimiento rige para caracoles terrestres de cualquier especie. Basta con tres ejemplares, pues éstos se multiplicarán. Además, la faena de mantener limpias las cajas con caracoles no es una bagatela, y aumenta en proporción al número de individuos.

Los caracoles comprados en un mercado le llegarán indudablemente en un estado latente llamado estivación, en el que caen cuando están calientes y secos. La estivación es la contrapartida de la hibernación, el período latente en que entran algunos animales durante los fríos meses de invierno.

Diversos observadores han informado que diferentes caracoles han sobrevivido en estivación durante cuatro o cinco años sin salir de su concha. Cuando el animal está en estivación no entra en actividad a menos que se le exponga a la humedad o se le perturbe. Cuando reciba los animales, manténgalos en un lugar seco hasta tenerles lista su caja.

Examine cuidadosamente cada ejemplar. Marque un distintivo en su concha con una laca coloreada, verbigracia un esmalte de uñas. Yo suelo anotar la marca de identificación en la parte superior de una página de un cuaderno. Peso cada ejemplar en un pesacartas y registro su peso en el cuaderno conjuntamente con su coloración y cualquiera marca distintiva que posea.

Los pequeños acuarios de vidrio que venden los comerciantes en peces tropicales son igualmente útiles. El nuevo recipiente debería lavarse con jabón y agua y enjuagarse bien. Coloque una delgada capa de hojas caídas, una hoja de lechuga, trozos de zanahoria y una o dos ramitas en el fondo, junto con un recipiente poco profundo que contenga de tres a cinco milímetros de agua y un segundo recipiente con tierra buena de la que se vende para maceteros. Estos materiales estimulan, al parecer, a los caracoles proporcionándoles un entorno diversificado en cuanto a sabores y olores.

El recipiente para la tierra puede ser un pequeño tiesto para flores o una caja rectangular de plástico. La capa de tierra debe tener un espesor de 2 cm por lo menos. Mantenga la tierra húmeda pero no anegada. Los animales pondrán sin duda sus huevos en ella.

Normalmente, los animales estarán más activos de noche, pero se puede invertir su ciclo de comportamiento para observarlos durante el día. Ilumine toda la noche el acuario mediante una bombilla brillante. Por la mañana, cubra el acuario con una tela opaca. Para observar los animales durante el día baje las persianas y quite la cubierta. Los caracoles pueden adaptarse a una inversión del día y la noche al cabo de una semana, aproximadamente. Si con esta técnica no logra estimularlos durante el día, inténtelo colocándolos en una fuente con agua de poca profundidad.

No mantenga sobrepoblado el acuario. Basta con tres caracoles. Si quiere seis, monte un segundo acuario. No introduzca en el recipiente ningún material que no pueda lavarse o eliminarse, excepto la tierra. Retire los caracoles y todos los materiales dos veces por semana y luego lave y enjuage minuciosamente el acuario para limpiarlo de limo y excremento. Arroje los alimentos no consumidos y reemplácelos con otros frescos y desechos vegetales naturales.

Los objetos naturales, tales como piedras y trozos de tiestos quebrados son adecuados como ornamentos del acuario. Debería incluirse siempre una concha de almeja o un hueso viejo. Se los irán comiendo poco a poco porque el caracol necesita una cantidad relativamente grande de calcio. Durante sus paseos por el campo puede juntar una haz de hojas caídas y ramitas, que guardará para el invierno.

Estimule a los caracoles sólo después

de tener listo el acuario. Todo lo que tiene que hacer es ponerlos sobre una superficie húmeda en su nuevo hogar. Tenga presente que está trabajando con caracoles terrestres y no acuáticos. No ponga más de tres milímetros de agua en el acuario, pues podrían ahogarse. Al cabo de unos 10 minutos, uno o más de los caracoles se activará y reptará subiendo por cualquier superficie cercana. Si alguno de ellos no ha dado señales de vida a las veinticuatro horas, quizás haya muerto. A los caracoles muertos se les puede identificar después de unos días por su fuerte hedor.

Observe frecuentemente a los animales durante el primer día después de que hayan salido de la estivación. Si en el interior del acuario el aire es húmedo, como debiera ser, los caracoles reptarán por los costados y pasarán, desafiando la gravedad, al anverso de la tapa. Usted debería poder ver al menos a uno de ellos levantar la parte anterior de su pie, su cabeza y sus tentáculos, separándolos de la tapa, retorciéndolos y repasando con su aparato bucal la abertura de su concha. El animal está comiendo su primer alimento después de ayunar durante muchos días o semanas, a saber, los restos de la película de mucus que cubre el cuerpo durante la estivación.

Los caracoles de tierra no son remilgados para comer. Aunque consumirán una amplia variedad de materias vegetales y animales, resulta entretenido descubrir sus preferencias alimentarias. En cierta ocasión ofrecí a cuatro caracoles un trozo de atún. Tres de ellos lo desdeñaron, pero el cuarto se lo comió entero. Evidentemente, los animales desarrollan preferencias individuales.

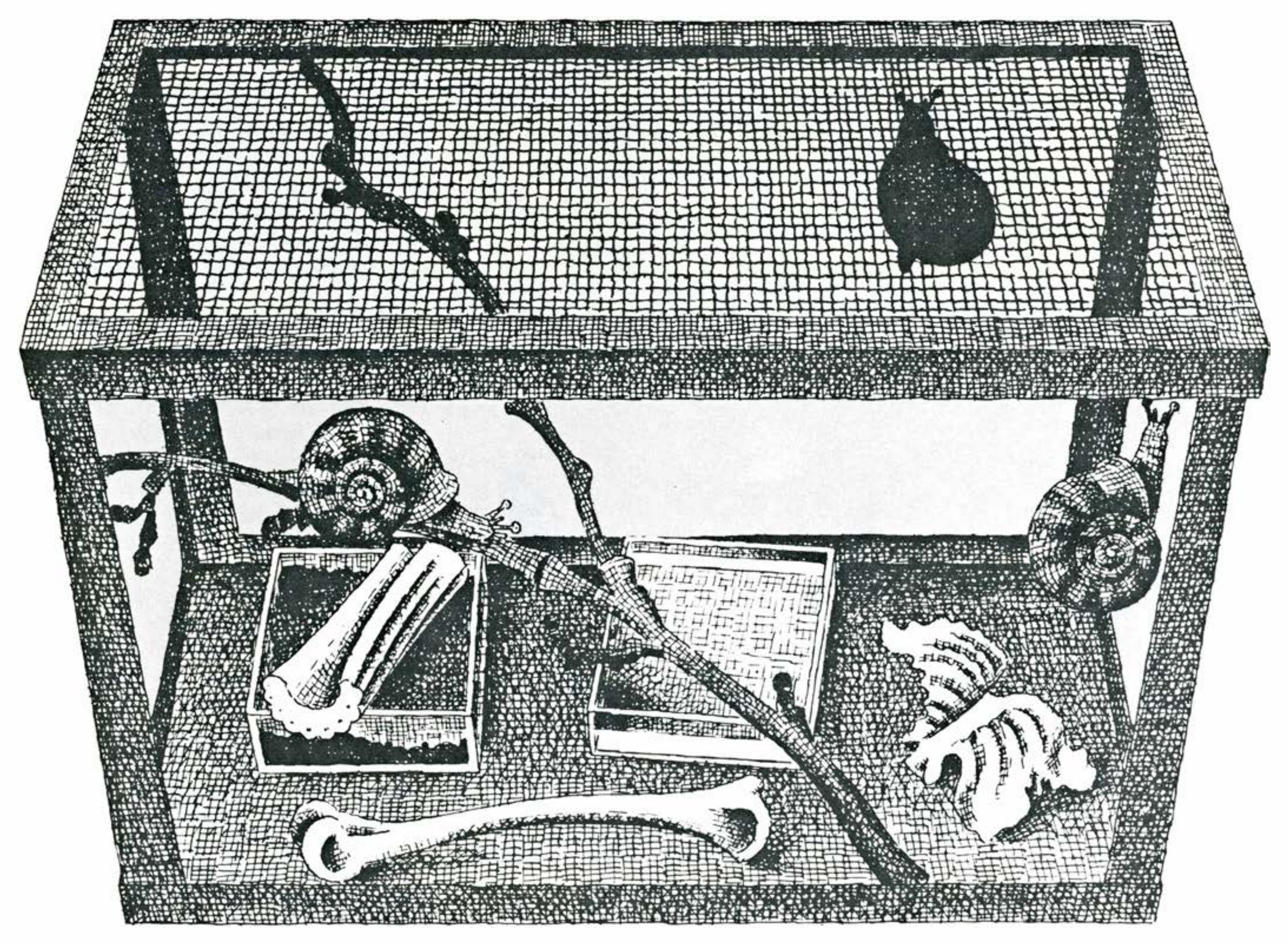
La dieta debería incluir cereales, verbigracia, copos de avena, afrecho y esos cereales envasados que se consumen en algunos países en el desayuno. Como ya he dicho, el calcio es esencial. Espolvoree unas pulgaradas de piedra caliza molida o de polvo de huesos sobre las verduras además de mantener una concha marina o un trozo de hueso.

El calcio es el material con que el caracol construye su concha, la cual crece durante toda la vida del animal desde el primer tabique. Los caracoles terrestres en su medio natural obtienen calcio de distintos materiales orgánicos. Así como el hombre ingiere calcio para sus huesos con la leche, los caracoles consumen los restos de otros animales. Si carecen de una adecuada provisión de calcio, los caracoles pueden entrar en estado de latencia

Los dos rasgos más prominentes de un caracol son su concha y su pie, que es la larga extensión del cuerpo sobre la cual se desliza. La boca está en la parte anterior del pie y puede observarse a través del vidrio o plástico cuando el animal se mueve. Cuando está abierta es un agujero circular y, al cerrarse, tiene forma de estría, como una T. La boca está algunas veces provista de maxilas.

El caracol conjuga las funciones de los dientes y la lengua en un solo órgano: la rádula, un músculo dentado, semejante a una lima, situado en el interior de la boca, que emplea como raspador para masticar el alimento. La rádula puede verse cuando la boca está abierta o cuando uno enfoca una luz fuerte a través de la cabeza por arriba y la observa desde abajo.

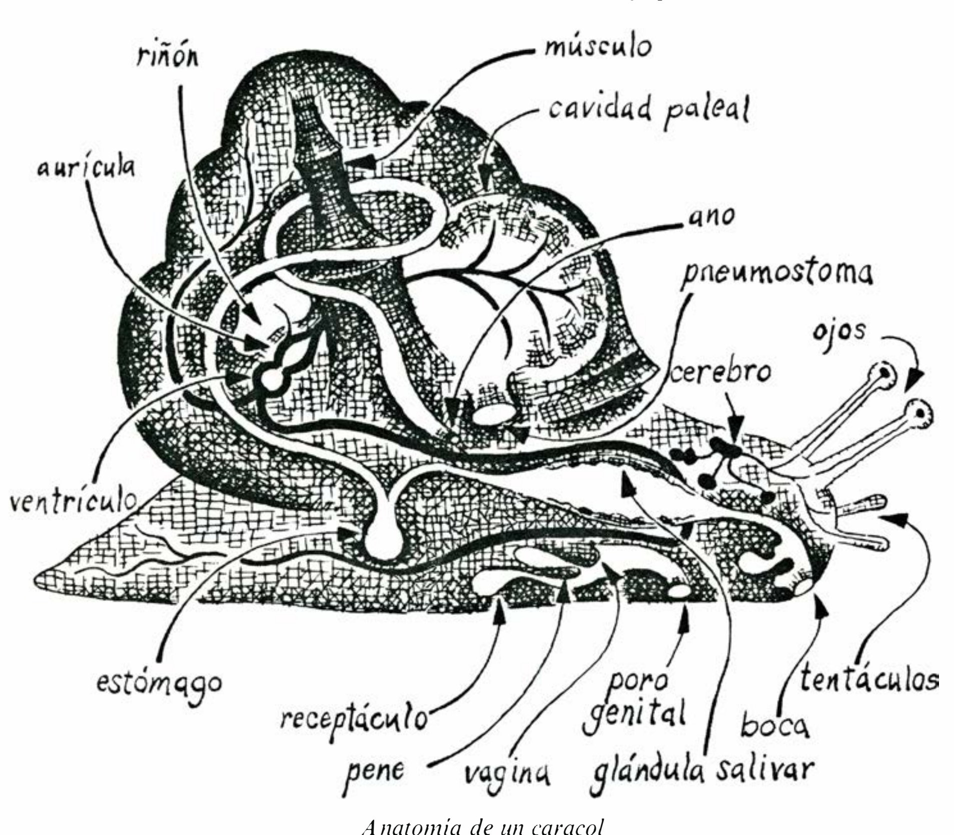
Mantenga silencio y escuche cuidadosamente cuando un caracol esté comien-



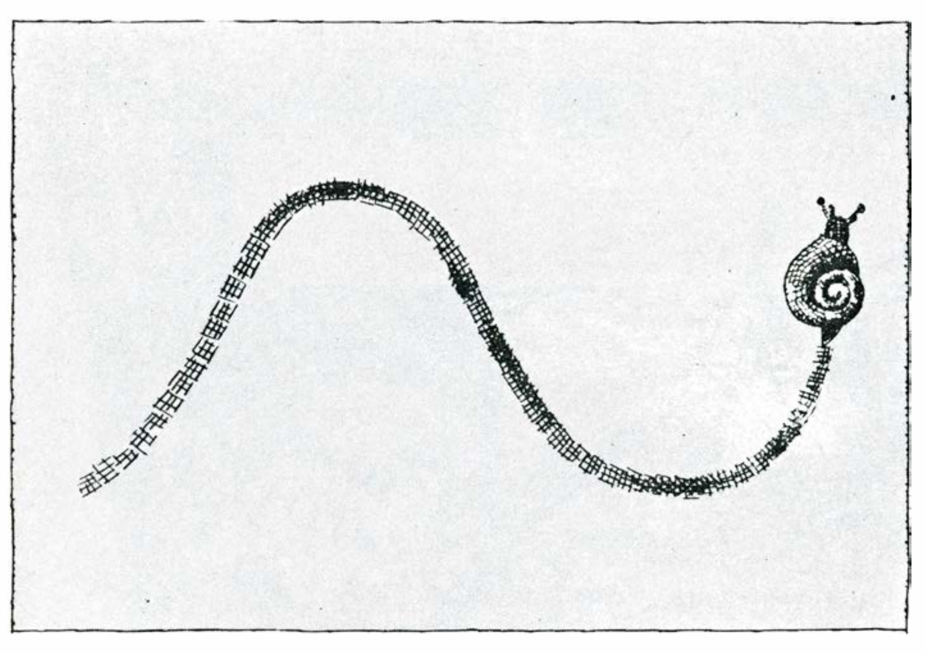
Dispositivo recomendado por Robert A. Brown para guardar caracoles

do materiales duros tales como lechuga fresca y zanahorias. Oirá un claro sonido de raspadura conforme la rádula desgasta el alimento. Se puede también sentir la rádula poniendo una delgada capa de miel o mermelada en la punta de un dedo y ofreciéndosela al caracol. El animal dejará el dedo limpio de esa capa usando la rádula. Es interesante anotar en el cuaderno las preferencias alimentarias de cada uno de los caracoles de un acuario.

El ano aparece en el costado derecho del cuerpo cuando el animal sale de su concha. Puede detectarse fácilmente cuando excreta, especialmente cuando ha comido alimentos muy coloreados. Observe en especial los órganos sensoriales de la cabeza, a saber los tentáculos anteriores y posteriores. Estos últimos son más largos. Al extremo de cada uno hay un ojo. Cuando se toca los tentáculos, se retraen invaginándose. Se les puede ver con detalle usando una luz brillante y una lupa. Los tentáculos retraídos aparecen por el procedimiento inverso de volverse hacia afuera.



Anatomía de un caracol


Tendencia a moverse hacia arriba conforme va cambiando la posición de una lámina vertical de vidrio

Observe el tejido oscuro y carnoso que circunda el borde de la abertura de la concha. Se le denomina manto. Entre otras funciones, segrega la concha. Forma asimismo una cavidad que, en el caracol de tierra, sirve como pulmón. Mirando hacia arriba, hacia la abertura de la concha, cuando el caracol ha salido de ella por completo, puede verse el gran agujero de ventilación que conduce a la cavidad del pulmón. Suele estar abierto.

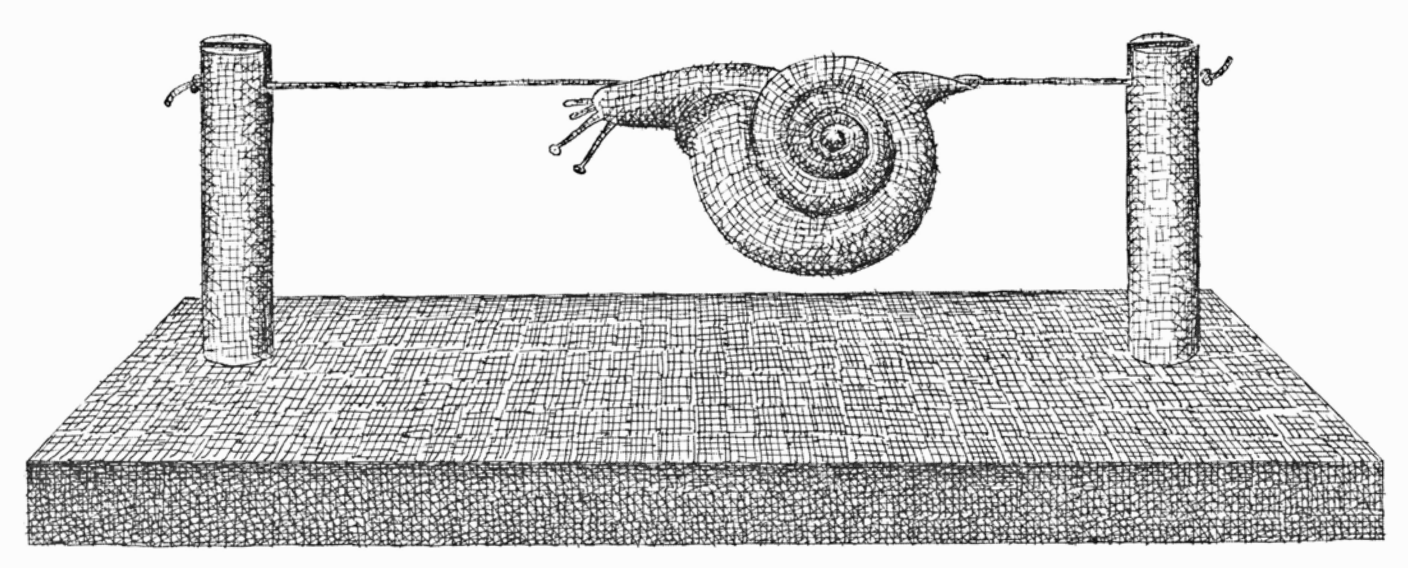
Los caracoles son hermafroditas. Cada ejemplar es al mismo tiempo macho y hembra. Los animales se juntan en un cortejo complicado y fácil de observar que culmina en la fertilización cruzada. Para aparearse, los caracoles se acarician uno a otro durante largo tiempo y luego se disparan uno al otro un dardo calcáreo antes de copular. El apareamiento se realiza cuando desde el lado derecho de la cabeza emerge el gran órgano sexual. El órgano sexual de cada caracol penetra en el órgano sexual de su pareja, y el esperma de cada animal fertiliza los huevos del otro.

Ambos caracoles depositan huevos del tamaño de una mota de polvo en la tierra. Cada vez que limpie el acuario, busque cuidadosamente los huevos. Perturbelos lo menos posible. Aunque son frágiles, pueden incubarse, y, si así ocurre, podrá montar un acuario separado para las crías.

Para investigar cómo se mueven los caracoles, ponga uno de ellos sobre una lámina de vidrio o plástico transparente. Mantenga vertical la lámina. El caracol se moverá hacia arriba. Haga girar lentamente en uno y otro sentido la lámina, manteniéndola vertical. El animal sentirá el cambio de la dirección de la gravedad y girará para mantener su curso ascendente.

Examine el movimiento del pie a través del vidrio. Unas ondas de contracción muscular parecerán desplazarse desde la parte trasera a la delantera del pie. Las ondas avanzan en sentido contrario al movimiento del caracol hacia adelante. De hecho, la parte del pie que se mueve hacia adelante no está en contacto con la superficie del vidrio y es, por consiguiente, menos notoria.

La huella prominente de mucus que aparece como estela del caracol es segregada por una glándula que podrá ver



Acrobacias de un caracol sobre una cuerda

inmediatamente detrás de la boca. El mucus hace posible el movimiento deslizante del caracol. Es fácil hacer una demostración de los talentos acrobáticos de estos animales. Por ejemplo, ponga uno de ellos sobre una cuerda o un hilo fino tendido horizontalmente entre dos soportes. El animal se trasladará a lo largo de la cuerda o hilo con igual facilidad con que lo hace sobre una superficie plana. Puede también deslizarse sobre superficies espinosas, como las púas de un cepillo, sin sufrir daño, ni, al parecer, molestia.

Mediante observación y unos pocos experimentos se puede comprobar que todas las partes del cuerpo del caracol sienten la luz, aunque el animal tenga ojos en el ápice de sus tentáculos más largos. La reacción del caracol ante la luz puede probarse de diversas maneras. Por ejemplo, instale una pequeña bombilla a cada lado del acuario. Encienda las bombillas de forma alterna, oscureciendo simultáneamente el lado no iluminado. El caracol se desplazará hacia el lado más oscuro.

El animal puede reaccionar ante la aproximación brusca de una pequeña sombra retirándose dentro de su concha. Pase la sombra de su mano sobre un caracol. El animal parece interpretar la sombra como una amenaza.

Unos investigadores que han extraído los ojos de los tentáculos mayores han llegado a la conclusión de que tienen poca importancia, puesto que la conducta de los animales privados de ellos no cambia significativamente. Parece que los ojos se fijan sobre objetos sólo a muy cortas distancias. Quizá vea ocasionalmente a un caracol doblar sus tentáculos oculares hacia abajo lo suficiente como para fijarlos sobre la superficie que lo sostiene.

¿Cómo reacciona un caracol ante los olores? Se puede probar el sentido del olfato colocando alimentos cocinados y calientes sobre una superficie humedecida. Observe los movimientos de búsqueda que realiza el animal, elevando la parte delantera de su pie y agitando sus tentáculos.

¿Cómo perciben los alimentos ¿Mediente el olfato o mediante la vista? Ponga dos o más pantallas opacas sobre la superficie, y el caracol entre ellas. Las pantallas pueden ser tiras de metal dobladas en forma de L. Observe el comportamiento del animal respecto de las pantallas. Luego ponga el alimento caliente detrás de las pantallas y observe las alteraciones del comportamiento.

¿Pueden distinguir estos animales entre alimentos de varias clases? ¿Muestran preferencia? Para saber la respuesta ponga una gota de vinagre en el camino de un caracol. Se echará a un lado. (No vierta vinagre sobre el animal). El caracol retraerá el borde de su pie que esté más cerca de la gota formando así un entrante. Este entrante se correrá desde el extremo anterior al posterior, a lo largo del pie, conforme el animal se vaya deslizando hacia adelante.

¿Oye un caracol? La respuesta es dudosa. El animal reaccionará ante ruidos fuertes y abruptos, quizás porque el aire hace vibrar la concha. El caracol tiene un sentido táctil finísimo. Los tentáculos se retraen con el más ligero toque. Por ejemplo, un solo hilo de telaraña que llegue flotando a tocar un tentáculo hará que este órgano se retraiga. Por otra parte, como ya he dicho, los objetos afilados no hieren el pie. Un caracol puede reptar sobre el filo de una hoja de afeitar sin cortarse.

Como es de suponer, el caracol reacciona vigorosamente ante la humedad. Después de limpiar el acuario, ponga un tabique de poca altura a lo ancho de la base, disponiendo una capa de grava seca de acuario a un lado y de grava húmeda al otro. ¿Hacia qué lado se mueve el caracol?

La reacción del animal ante la gravedad está evidentemente relacionada con su sentido del tacto. Basándose en esta reacción, podrá construir un divertido aparato. Monte un columpio de balancín en miniatura cementando la parte media de un portaobjetos de vidrio para microscopía a un trozo de cuerda que se tense a unos dos centímetros por encima de una mesa. Ponga el caracol en el portaobjetos, por el reverso de éste. El animal reptará ascendiendo por la plaquita, la cual se inclinará hacia abajo cuando aquél cruce el punto de equilibrio. El caracol dará entonces la vuelta y avanzará en dirección opuesta hasta que el columpio vuelva a inclinarse. Limite el experimento a unos pocos ciclos.

¿A qué paso marcha un caracol? La velocidad máxima a la que se mueve un caracol parece depender de varios factores. Los caracoles tienden a moverse cuesta arriba, hacia la sombra, hacia alimentos de su predilección y hacia la humedad. Algunos caracoles quizá respondan más vigorosamente que otros a diversas combinaciones de estos estímulos, por lo cual la velocidad varía según los individuos y las circunstancias. Me parece interesante cronometrar el avance de cada animal a lo largo de una distancia media de, por ejemplo, 20 centímetros cuando se mueve apartándose de la luz, ascendiendo y hacia la humedad. Registrando las preferencias de distintos caracoles se puede calcular las probabilidades para fijar los "handicap" en una carrera de caracoles.

Libros

La peripecia del lingüista que osa meterse a filósofo

Juan José Acero

La teoría de las ideas innatas en Chomsky, José Hierro S.- Pescador. Editorial Labor, Barcelona 1976. En esta obra, a la que me referiré en forma abreviada como TIICH, el autor examina los presupuestos psicológicos y epistemológicos que presiden el desarrollo de la lingüística generativo-transformacional, con particular incidencia en el credo que N. Chomsky profesa al respecto.

La presente reseña se desarrolla de la siguiente forma. En primer lugar, indicaré cuál es el contexto teórico al que TIICH pertenece y me haré eco de sus conclusiones principales. En segundo lugar, procederé a comentar algunas de las ideas allí expuestas sobre la evaluación que de la posición de Chomsky lleva a cabo J. Hierro. Finalmente, me referiré a un par de resultados debidos a S. Peters y R. W. Ritchie, familiares para los profesionales de la lingüística, y que tienen una estrecha relación con el contenido de TIICH.

Hoy día se ha convertido en un tópico la exigencia de que la teoría lingüística sea mentalista, es decir, que se le exige tener como objeto propio al hablante-oyente ideal y al conocimiento que éste tiene de su lengua. Esta afirmación es mucho más fuerte que la de limitarse a declarar que el estudio de las lenguas naturales ha de proporcionar información esencial acerca de las propiedades de la mente humana, pues esta idea es compatible con otra concepción de la teoría lingüística que persiga simplemente caracterizar la actuación del hablante-oyente. La posición mentalista impone el requisito de que la investigación de la actuación no puede ir más lejos de lo que el claro entendimiento de la competencia subyacente le permita. Para muchos teóricos (pertenecientes a campos muy distintos: bien el de la psicología, bien el de la filosofía, e incluso el de la lingüística), este ideario ha resultado muy difícil de roer. Su conclusión ha sido la controversia larga, y a menudo coloreada por ariscos exabruptos, que durante los últimos casi veinte años ha distinguido a la gran mayoría de estudios de metodología lingüística. Entre unos y otros han acumulado suficiente leña seca como para que la hoguera de la discusión no tenga visos de finalizar en un futuro inmediato. TIICH tiene aquí su lugar.

En el primer capítulo de su estudio ("Filosofía, psicología y lingüística en Chomsky"), el autor sienta con brevedad y precisión las bases de lo que será su análisis posterior. Toca temas como los de los niveles de adecuación de la teoría lingüística, los componentes centrales de la gramática generativa de una lengua natural, etc., para acabar formulando la pregunta principal que es necesario responder desde la perspectiva mentalista de Chomsky: ¿Qué aporta el sujeto al aprendizaje del lenguaje humano? El transcurso que va desde la concepción de la teoría lingüística al del aprendizaje del lenguaje humano tiene su razón de ser en el hecho de que el problema de la adecuación explicativa de la teoría lingüística es, en resumidas cuentas, el de construir una teoría de la adquisición del lenguaje. La dificultad aquí es doble: por un lado, es preciso caracterizar la noción del lenguaje biológicamente posible; por otro, hace falta proporcionar un procedimiento evaluativo en términos del cual seleccionar, de entre una diversidad de gramáticas posibles de una lengua dada, la que describe adecuadamente la competencia lingüística de su hablante-oyente ideal. Este programa tiene únicamente sentido en tanto en cuanto se postula, al mismo tiempo, que el sujeto que aprende su lengua nativa se halla en posesión de (o conoce) las propiedades determinantes de todo lenguaje biológicamente posible y los mecanismos necesarios para extraer de aquí la gramática de su lengua particular sobre la base de un reducido número de datos que le vienen dados por su entorno social. Típico de la actitud mentalista en lingüística es, pues, acreditar a todo sujeto con un conocimiento de contenidos innatos altamente abstractos.

Lo que distingue a este peculiar tipo de conocimiento es el objetivo del capítulo segundo ("Conocimiento del lenguaje y competencia lingüística"). Este capítulo es una versión ampliada de un previo ensayo del autor, titulado "Conocimiento, creencia y competencia lingüística". El núcleo de ambos lo constituye la cuestión de si el conocimiento del lenguaje al que se apela al hablar de la competencia lingüística de un hablante-oyente (ideal) o bien es conocimiento por familiaridad o bien una forma de *saber-que* o bien de *saber-cómo*. Dejando a un lado otros detalles de interés, la conclusión de Hierro es que "mi interpretación no ha superado la dicotomía saber-que/saber-cómo, y que lo único que ha mostrado es que el conocimiento del lenguaje es un conocimiento complejo que participa de ambas formas. Un saber-que que consiste en la aplicación inconsciente de reglas, y un saber-cómo que consiste en distinguir ciertas oraciones como correctas y otras como incorrectas" (p. 32-33). En el capítulo tercero, se expone el objeto de esta forma compleja de conocimiento ("La adquisición del lenguaje: los universales lingüísticos"), así como el tema de los márgenes dentro de los cuales resulta lícito hablar de un mecanismo innato poseído por todo usuario de una lengua.

Uno de los aspectos más interesantes del TIICH es su discusión del supuesto entronque entre el modelo chomskiano de adquisición de la lengua y la tradición filosófica racionalista; es decir, del carácter racionalista de la teoría lingüística. Todo el capítulo cuarto ("¿Ideas? ¿Innatas?") está dedicado a rechazar la filiación repetidamente promocionada por Chomsky. Esto se hace presentando la doctrina general cartesiana de las ideas innatas, considerando diversos matices según los autores que las propugnaron y pasando revista a los posibles paralelismos existentes entre aquélla y la teoría de los universales lingüísticos. Los autores examinados son Descartes, Herbert de Canterbury, Cudworth y Leibniz.

En el capítulo quinto ("Críticas") Hierro enumera las principales críticas dirigidas a la hipótesis chomskiana de las ideas innatas: las de H. Putnam, N. Goodman y W. Quine, y toma partido decididamente por el modelo de apren-

dizaje del lenguaje esbozado por este último en el influyente libro *Palabra y Objeto* (Editorial Labor, 1968) y matizado en una diversidad de ensayos posteriores. La viabilidad de esta alternativa, que no difiere de la actitud de Chomsky en cuanto a la necesidad de postular algún mecanismo innato, sino en cuanto al tipo de mecanismo requerido, ocupa el capítulo final del estudio de Hierro, titulado "Alternativas". En mi opinión, el punto a destacar aquí es la distinción entre innatismo biológico e innatismo filosófico. (El contexto en que se presenta esta dicotomía es el del empleo de la teoría biológica del aprendizaje del lenguaje de E. Lenneberg como fiel de la balanza en la que sopesar las doctrinas de Chomsky y de Quine.) Para Hierro, el paso de uno a otro es el paso de lo controlable y lo interesante a lo incontrolable y absurdo. Chomsky, cree el autor, da este paso. Su caso "ilustra una vez más la peripecia del científico que se mete a filósofo" (p. 123).

Desde sus páginas iniciales hasta las postreras, TIICH es una obra abierta y radicalmente crítica de la doctrina chomskiana de las ideas innatas, cuyo tono general puede resumirse, en palabras del autor, diciendo que las afirmaciones de Chomsky resultan "desproporcionadas en relación con los datos de que disponemos" (p. 135). Este comentario me parece del todo justo. Así, por ejemplo, en el ensayo de Th. Bever *"The Psychology of Language and Structuralist Investigations of Nativism"* se da cumplida cuenta del hecho de que los estudios experimentables sobre la realidad psicológica de la competencia lingüística no demuestran que las operaciones transformacionales que asignan una forma interior a una forma externa de una oración se correspondan con operaciones psicológicas reales. Y de que esto abre la posibilidad de que esa comprensión del lenguaje por parte de sus usuarios nativos involucre estructuras mentales que sean independientes de las gramaticales. Tanto el resultado experimental citado, como la hipótesis que de él parece desprenderse, nos autorizan a considerar con alguna desconfianza la atribución (que hace Chomsky) de estructuras innatas al aprendiz de una lengua, en particular las que se refieren a la significación psicológica de ciertos universales formales. Vale la pena destacar en este contexto que la anterior conclusión es compatible con el modelo de competencia lingüística esbozado en TIICH, III.5 ("Razones y sinrazones del modelo chomskiano"), basado en una perspectiva genética o dinámica del aprendizaje del lenguaje. El autor llega a decir que entonces "difícilmente habrá que recurrir a postular el innatismo de los universales lingüísticos" (p. 60). Esta afirmación es desde luego muy fuerte y puede resultar útil para contrarrestar salidas de Chomsky con un sabor típicamente polémico –he aquí una de su reciente ensayo *Problems and Misteries in the study of Human Language:* "... en su mayor parte, el lenguaje realmente no se enseña. Mejor [sería decir que] se aprende"–, pero, en general, no creo que haga justicia al resto de la posición de J. Hierro. Este mismo matiza a renglón seguido cuando dice que "un modelo de este tipo... reduce considerablemente, en mi opinión, la plausibilidad de los argumentos chomskianos a favor del innatismo de los universales lingüísticos" (p. 60). La cuestión está en delimitar el grado de esta reducción. No he podido encontrar en el libro de Hierro pasaje alguno incompatible con el principio metodológico fundamental del mentalismo de la teoría lingüística: a saber, la tesis de que la posibilidad de una teoría de la actuación reside en una teoría de la competencia lingüística.

El autor insiste en que la actitud es razonable en tanto en cuanto tenga consecuencias empíricas, de forma que éstas resulten el árbitro final de todo el problema. Esta forma débil, si así se quiere, de caracterizar la posición mentalista, frente a la de un innatismo de estructuras lingüísticas, alocado por irrestricto, también es para mí una forma razonable de enfocar el estudio de las lenguas naturales. Sin embargo, la investigación experimental ha puesto de manifiesto que este planteamiento tiene una base empírica indudable, porque a partir de él se demuestra la relevancia psicológica de la forma interior que una gramática transformacional adscribe a una oración en tres tipos de problemas empíricos: los de la memoria, los de la ambigüedad y los de la percepción lingüística. Otro género de estudios experimentales –los concernientes a la elaboración de gramáticas probabilísticas para fragmentos de lengua naturales, acometida por P. Suppes y otros colaboradores suyos– abundan en el mismo presupuesto. Todo esto sugiere que la aceptación de la doctrina chomskiana de las ideas innatas debe condicionarse a investigaciones ulteriores de una diversidad de tipos, entre los que destacan los referentes al lugar de las denominadas estrategias perceptivas *(perceptual strategies)* en la distinción entre competencia y actuación. La misma preocupación recorre TIICH de punta a cabo, por lo que constituye una guía muy útil para determinar los problemas a los cuales enfrentarse y resolver. Por lo demás, y éste es un inciso mío, no creo que la posición de Hierro sea irreconciliable con la de Chomsky.

A título personal quiero poner especial énfasis en el contenido de los capítulos segundo y cuarto. El lector podrá, por su parte, discrepar de esta valoración mía. Sin prejuzgar ahora nada sobre el asunto, el primero de estos dos capítulos me parece un ejemplo de análisis conceptual de gran calidad; el segundo, la elaboración de un argumento que deja en su justo punto la presunta fuerza del entronque entre la doctrina chomskiana de la adquisición del lenguaje y la tradición del pensamiento racionalista. Recientes comentarios tanto de Chomsky a solas (en el ensayo que he mencionado más arriba), así como junto con Katz (en la réplica que ambos han dado a Cooper) acerca de este tema, le permitirían a Hierro llegar a un acuerdo razonable con ambos lingüistas. Esta es mi creencia, quizás un tanto optimista.

Sin embargo, voy a llevar mi optimismo algo más lejos y a atreverme a decir que tampoco aprecio, en la elaboración que hace Hierro de la disputa entre Chomsky y Quine, razones serias para sostener a ultranza el antagonismo que el primero de los tres contempla en la posición respectiva de los últimos. La clave de la controversia reside en el concepto de disposición, y en TIICH, se nos dice que "pensar que las disposiciones pertenecen a la ejecución, al habla, es no haber entendido nada de lo que Quine quiere decir" (p. 115). Y que "'disposiciones' está al nivel de 'competencia'" (p. 115). Este inciso me resulta del todo iluminador, aunque me temo que no conste explícitamente así en ninguno de los escritos de Quine. Aunque Hierro no lo manifiesta, esta reconstrucción del empirismo de Quine sería del tipo de las que Chomsky acogería con toda simpatía. Contra ella, en mi opinión, Chomsky no tendría absolutamente nada que objetar, al menos si juzgamos a través de su reciente escrito antes citado. Así, al tratar del principio de aplicación cíclica de las reglas transformacionales, ha afirmado que sería posible (es decir, legítimo) concebirlo como una disposición a adquirir una cierta competencia (una cierta estructura cognitiva, una gramática, un conocimiento del lenguaje). Habiendo adquirido esta estructura cognitiva, una persona puede estar dispuesta a hablar de ciertas formas. Ahora bien, la aclaración anterior de Hierro

está libre de las críticas de Chomsky porque parte de, y presupone, la exigencia que éste postula: la distinción entre competencia y actuación lingüística. Y es justamente el olvido de este punto lo que promueve las réplicas de los teóricos orientados hacia el mentalismo. Chomsky añade, además, machaconamente, que esta manera de entender el concepto de disposición se aparta del ideario conductista, estrictamente interpretado; a lo que se le suele responder que tal diagnóstico no hace justicia a los requisitos metodológicos básicos de la investigación científica. Semejante derivación de la polémica amenaza en convertirla en una disputa por una recompensa pírrica: ¿por qué toda remodelación del conductismo en psicolingüística tiene que estar abocada a llevar sobre sí la impronta skinneriana más como un estigma que como un distintivo? ¿Por qué, desde el otro lado de la barrera, el discurso sobre ideas innatas está *en principio* viciado?

En su reseña de la edición ampliada de *El lenguaje y la mente,* G. Harman opina que la polémica en torno a los presupuestos metodológicos y psicológicos de la teoría lingüística perderá su razón de ser cuando los malentendidos originados por las partes en disputa queden al descubierto. Yo comparto esa idea. Buen número de ellos –con un énfasis especial en los suscitados por los lingüistas metidos a filósofos– están claramente denunciados en la obra de J. Hierro. En mi opinión, ahí reside su principal mérito.

El menos exigente de los tres criterios por los que se debe juzgar la adecuación empírica de una gramática es el denominado criterio observacional: una gramática es observacionalmente adecuada cuando predice qué secuencias de fonemas de una lengua son oraciones (gramaticales) de ésta y cuáles no lo son. Por tanto, es adecuada si presenta con justeza los datos primarios observados por el lingüista. Este criterio, en tanto que se asume para reflejar en la teoría ciertas habilidades de los usuarios de una lengua, conduce a la hipótesis de que sus oraciones constituyentes se organizan bajo la forma de un conjunto recursivo (decidible). Esta hipótesis, formulada explícitamente por H. Putnam en 1961, no es fácil de contrastar. Para conseguirlo, hay que marcarse ciertos límites y seguir pautas aparentemente restrictivas como indicadores de la evidencia valedera. He aquí tres ejemplos: si en una situación dada un hablante considera gramatical una oración, el lingüista la aceptará como tal en términos absolutos; el lingüista se atendrá sólo a rasgos puramente formales, o estructurales, y no materiales, o de contenido, con respecto a cada oración que considere; como árbitros últimos, el lingüista apelará a sujetos pertenecientes a un mismo grupo dialectal, aunque quede a su arbitrio qué grupo escoger. Dejando ahora a un lado el hecho de que cada una de estas pautas es en sí misma problemática, durante algunos años –todavía hoy en día, me atrevería a decir– se pensó que la hipótesis de la recursividad del conjunto de las oraciones de una lengua tenía implicaciones contrastadoras de orden psicolingüístico lo suficientemente importantes como para que su suerte estuviese ligada a la de la concepción de la teoría gramatical misma.

Sin embargo, a finales de la década de los años sesenta, S. Peters y R. Ritchie alcanzaron un resultado sorprendente operando con el modelo de análisis gramatical propuesto por J. Katz y P. Postal en 1964, que Chomsky hizo también suyo en *Aspectos de la teoría de la sintaxis:* las gramáticas transformacionales elaboradas no generaban conjuntos recursivos de oraciones, sino conjuntos recursivamente enumerables de oraciones. Lo que, dicho a grandes rasgos, significaba que no eran observacionalmente adecuadas, vista la cuestión desde la perspectiva de factores restrictivos como los mencionados en el párrafo anterior.

Este resultado determinó en gran medida la investigación subsiguiente, pues confiando pese a todo en la conveniencia de contar con el mencionado criterio de adecuación observacional, lo que quedaba en la picota eran precisamente las pautas que determinaban qué género de datos constituían evidencia valedera. Así, buena parte de la viva polémica entre la semántica interpretativista y la semántica generativista puede comprenderse como una disputa relativa al segundo de los tres ejemplos anteriores. Pero incluso en este respecto, en el que la investigación reciente ha sido intensísima y en el que los aires de las modas parecen haber cambiado de dirección, carecemos de resultados concluyentes: una diferencia de criterios acerca de si el lingüista tiene que limitarse o no a tomar nota de factores puramente formales o estructurales en una oración no conlleva forzosamente una diferencia en las implicaciones contrastadoras de la hipótesis inicial. Basta con recordar a título de ejemplo revelador la cuestión de la naturaleza de la base de una gramática transformacional y un segundo resultado de Peters y Ritchie: la existencia, con respecto a toda lengua, de una gramática descriptivamente adecuada de ella dotada de una base universal.

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

LA IMPORTACION DE GAS NATURAL LICUADO

LNG INFORMATION BOOK, 1973. Editado por el LNG Information Book Task Group of the Liquefied Natural Gas Committee. American Gas Association, 1973.

LIQUEFIED NATURAL GAS. W. L. Lom. John Wiley & Sons, 1974.

TECHNOLOGY AND CURRENT PRACTICES FOR PROCESSING, TRANSFERRING, AND STORING LIQUEFIED NATURAL GAS. Arthur D. Little, Inc., para la Office of Pipeline Safety, 1974.

LIQUEFIED NATURAL GASES: VIEWS AND PRACTICES, POLICY AND SAFETY. U.S. Coast Guard, 1976.

LA COLISION ENTRE LA INDIA Y EURASIA

GEOLOGY OF THE HIMALAYAS. Augusto Gansser. Interscience Publishers, 1964.

TIBETAN, VARISCAN, AND PRECAMBRIAN BASEMENT REACTIVATION: PRODUCTS OF CONTINENTAL COLLISION. John F. Dewey y Kevin C. A. Burke en *The Journal of Geology,* vol. 81, n.º 6, págs. 683-692; noviembre, 1973.

CENOZOIC TECTONICS OF ASIA: EFFECTS OF A CONTINENTAL COLLISION. Peter Molnar y Paul Tapponnier en *Science,* vol. 189, n.º 4201, págs. 419-426; agosto, 8, 1975.

THE SUBDUCTION OF THE LITHOSPHERE. M. Nafi Toksoz en *Scientific American,* vol. 233, n.º 5; noviembre, 1975.

SUR LE MÉCANISME DE FORMATION DE LA SCHISTOSITÉ DANS L'HIMALAYA. Maurice Mattauer en *Earth and Planetary Science Letters,* vol. 28, n.º 2, págs. 144-154; diciembre, 1975.

SLIP-LINE FIELD THEORY AND LARGE-SCALE CONTINENTAL TECTONICS. Paul Tapponnier y Peter Molnar en *Nature,* vol. 264, n.º 5584, págs. 319-324; noviembre 25, 1976.

EL INTERFERON

INTERFERON. D. C. Burke en *Defence and Recognitions,* dirigido por R. R. Porter. University Park Press, 1973.

INTERFERONS AND INTERFERON INDUCERS. Dirigido por N. B. Finter. American Elsevier Publishing Co., Inc., 1973.

SELECTIVE INHIBITORS OF VIRAL FUNCTIONS. Dirigido por William A. Carter. CRC Press, 1973.

THE MECHANISM OF ACTION OF INTERFERON. David H. Metz en *Cell,* vol. 6, n.º 4, págs. 429-439; diciembre, 1975.

ALGORITMOS

COMPUTER SCIENCE AND ITS RELATION TO MATHEMATICS. Donald E. Knuth en *The American Mathematical Monthly,* vol. 81, n.º 4, págs. 323-343; abril, 1974.

ORDERED HASH TABLES. O. Amble y D. E. Knuth en *The Computer Journal,* vol. 17, págs. 135-142; mayo, 1974.

BINARY SEARCH TREES AND FILE ORGANIZATION. J. Nievergelt en *Computing Surveys,* vol. 6, n.º 3, págs. 195-207; septiembre, 1974.

THE COMPLEXITY OF SEARCHING AN ORDERED RANDOM TABLE. Andrew C. Yao y F. Frances Yao en *Symposium on Foundations of Computer Science.* IEEE Computer Society's Technical Committee, IEEE Computer Society, 1976.

BIOCRISTALES

SKELETON FORMATION OF SEA URCHIN LARVAE, IV: CORRELATION IN SHAPE OF SPICULE AND MATRIX. Kayo Okazaki en *Embryologia,* vol. 17, n.º 1, págs. 21-28; enero, 1962.

EXPERIMENTAL EMBRYOLOGY OF ECHINODERMS. Sven Hörstadius. Clarendon Press, 1973.

SPICULE FORMATION BY ISOLATED MICROMERES OF THE SEA URCHIN EMBRYO. Kayo Okazaki en *American Zoologist,* vol. 15, n.º 3, págs. 567-581; verano, 1975.

LOS ACOMPAÑANTES DE LAS ESTRELLAS DE TIPO SOLAR

DOUBLE STARS. Peter van de Kamp en *Publications of the Astronomical Society of the Pacific,* vol. 73, n.º 435, págs. 389-409; diciembre, 1961.

INTELLIGENT LIFE IN THE UNIVERSE. I. S. Shklovskii y Carl Sagan. Holden-Day, Inc., 1966.

BINARY AND MULTIPLE SYSTEMS OF STARS. Alan H. Batten, Pergamon Press, 1973.

EVOLUCION CROMOSOMICA DE LOS PRIMATES

COMPARATIVE MAMMALIAN CYTOGENETICS. Kurt Benirschke, ed. Springer, N. Y., 1969.

COMPARATIVE GENETICS IN MONKEYS, APES AND MAN. Brunetto Chiarelli, ed. Academic Press, N. Y., 1971.

ANIMAL CYTOGENETICS. MAMMALIA II: PLACENTALIA 5, PRIMATES. José Egozcue, Gebrüder Borntraeger, Stuttgart, 1975.

TEORIA DEL ARCO IRIS

LIGHT SCATTERING BY SMALL PARTICLES. H. C. Van de Hulst. John Wiley & Sons. 1957.

LIGHT AND COLOUR IN THE OPEN AIR. M. Minnaert, trad. al inglés por H. M. Kremer-Priest y revisado por K. E. Brian Jay. G. Bell and Sons Ltd. 1959.

THE RAINBOW: FROM MYTH TO MATHEMATICS. Carl B. Boyer. Thomas Yoseloff, 1959.

INTRODUCTION TO METEOROLOGICAL OPTICS. R. A. R. Tricker. American Elsevier Publishing Co. Inc., 1970.

THEORY OF THE RAINBOW. V. Khare y H. M. Nussenzveig en *Physical Review Letters,* vol. 33, n.º 16, págs. 976-980; octubre 14, 1974.

TALLER Y LABORATORIO

HYMAN SERIES IN INVERTEBRATE BIOLOGY: ZOOLOGICAL SCIENCES. vol. 6: MOLLUSCA I. Libbie H. Hyman. McGraw-Hill Book Company, 1967.

A SNAIL'S PACE. Lilo Hess. Charles Scribner's Sons, 1974.